



## Goal arguments of *ir* 'to go' and *chegar* 'to arrive' in three African varieties of Portuguese

**Tjerk Hagemeijer**, School of Arts and Humanities, University of Lisbon, PT, [t.hagemeijer@letras.ulisboa.pt](mailto:t.hagemeijer@letras.ulisboa.pt)

**António Leal**, University of Porto, Centre of Linguistics of the University of Porto, PT, [jleal@letras.up.pt](mailto:jleal@letras.up.pt)

**Raquel Madureira**, School of Arts and Humanities, University of Lisbon, PT, [raquelmadureira@edu.ulisboa.pt](mailto:raquelmadureira@edu.ulisboa.pt)

**João Cordeiro**, University of Beira Interior, HULTIG, LIAAD, INESC TEC, PT, [jpaulo@di.ubi.pt](mailto:jpaulo@di.ubi.pt)

---

In research on African varieties of Portuguese, especially Angolan and Mozambican Portuguese, it is often referred that Goal arguments of verbs of movement show a tendency to be headed by locative preposition *em* 'in', contrasting with the use of *a* 'to' and *para* 'to, toward' in European Portuguese. Language contact is generally considered the primary factor with respect to the use of the noncanonical pattern. Using spoken corpora of the urban varieties of Angolan, Mozambican and Santomean Portuguese, this paper develops a case-study of Goal arguments that occur with two frequently used verbs of inherently directed motion, *ir* 'to go' and *chegar* 'to arrive', to assess the contact-induced hypothesis and to explore alternative, semantic-based hypotheses. Overall, a cross-comparison of the varieties at stake and their main contact languages shows that the role of language contact is limited at best. A semantic analysis of *em*, on the other hand, shows promising results, since the occurrence of this preposition is favored by NPs that denote an entity with well-defined boundaries and is sensitive to the lexical semantics of the verbs, i.e., whether the verb describes a durative or non-durative change of place, and to the type of eventuality described by the predication.

---



## 1. Introduction

The literature on postcolonial, nativizing African varieties of Portuguese (AVPs) spoken in Angola (AP), Mozambique (MP), and São Tomé and Príncipe (STP) refers a more widespread and non-standard use of locative preposition *em* ‘in’ selecting the Goal argument of directed motion verbs (e.g., Adriano, 2014; Avelar, 2017; Chavagne, 2005; Mingas, 2000; P. Gonçalves & Chimbutane 2004; P. Gonçalves, 2010; R. Gonçalves, 2010), which contrasts with the use of the standard directional prepositions *a* ‘to’ and *para* ‘to, toward’ used in European Portuguese (EP), the official standard in the former Portuguese colonies in Africa. In the case of AP and MP, the role of language contact with Bantu languages has been at the forefront of the descriptive and theoretical explanations regarding these patterns (e.g., Avelar & Álvarez López, 2018; Marques, 1985; Mingas, 2000; P. Gonçalves & Chimbutane, 2004).

Since these previous claims regarding the expression of Goal arguments are typically not based on quantified data nor on comparisons across AVPs, we aim to provide a corpus-based assessment of the expression of Goal arguments of two high frequency verbs of inherently directed motion (cf. Levin, 1993), *ir* ‘to go’ and *chegar* ‘to arrive’, in the three varieties at stake, in order to discuss the factors that drive language variation in this domain. Since language contact has been a primary explanation, the point of departure will be a cross-comparative approach between the three AVPs and their main contact languages in the urban areas where the corpora were collected: Bantu languages Kimbundu (KB) for AP (Luanda) and Changana (CH) for MP (Maputo), and creole language Forro (Santome) for STP (São Tomé). Since it will be shown that the contact-induced hypothesis does not offer an encompassing explanation, we will shift our focus to an analysis that takes into account semantic features of the constructions under discussion.

The paper is organized as follows. Section 2 sets the stage with respect to the three varieties at stake and the corpus-based data used in this paper; section 3 discusses the expression of Goal arguments in the three AVPs with motion verbs *ir* ‘to go’ and *chegar* ‘to arrive’ based on quantified data, showing the patterns of variation within and across these varieties; section 4 provides an overview of the properties of Goal arguments in the main contact languages of the AVPs, namely Kimbundu, Changana, and Forro; section 5 relates and discusses the findings of sections 3 and 4 from the perspective of language contact, arguing that this factor lacks explanatory power to account for the observed patterns; finally, in section 6, we offer an alternative explanation for the data at stake that takes into consideration the semantics of the predications of the verbs *ir* and *chegar* and of the complements of the prepositions, focusing on features such as durativity and the type of location.

## 2. Background and methodology

Portuguese in Africa has been historically acquired as an L2, a phenomenon that essentially has its roots in the late 19<sup>th</sup> and 20<sup>th</sup> century when the former Portuguese colonies were

effectively colonized (e.g., Hagemeyer, 2016; P. Gonçalves, 2013). In addition, Portuguese in Angola, Mozambique, and São Tomé and Príncipe also shows an increasing tendency toward nativization since their independences in 1975, boosted by the choice of Portuguese as the only official language and its widespread democratization in all these postcolonial societies, hereby reinforcing its role as a lingua franca. Portuguese is by far the most spoken language in São Tomé and Príncipe (98,4%), whereas the predominant creole, Forro, is nowadays only spoken by 36,2% of the population (INE, 2013); in Angola, 71,15% of the population indicates that Portuguese is the most spoken language at home, especially in urban areas, whereas Kimbundu, one of the main Bantu languages which traditionally competes with Portuguese in the capital Luanda, is only spoken by 7,82% (INE, 2016); in Mozambique, Portuguese is spoken by around half of the population, which includes 16,6% of L1 speakers (INE, 2019), also with prevalence in urban areas. Despite the lack of direct information on L1 and L2, it can be inferred that Portuguese in Luanda and São Tomé is generally L1 and/or the primary language, with increasing monolingualism, especially among the younger generations, which make up a substantial part of the total population.

This case study is based on spoken urban corpora of MP, AP, and STP that were prepared within the project *Possession and Location: microvariation in African varieties of Portuguese* (PALMA) (Hagemeyer et al., 2021; Miguel et al., 2021; R. Gonçalves et al., 2021). The corpora were collected in the capitals Maputo, Luanda, and São Tomé between 2008 and 2020, by several researchers of the Center of Linguistics of the University of Lisbon, as well as doctoral students, and were as much as possible balanced according to level of education, age, and gender (cf. Hagemeyer et al., 2022). The interviews that integrate the corpora are predominantly semi-structured. Portuguese is the L1 or primary language of the majority of the informants who contributed to the corpora, especially in the case of urban AP and STP. **Table 1** summarizes the basic information of the corpora used.

	<b>Interviews</b>	<b>Hours</b>	<b>Tokens</b>	<b>Years of recording</b>
<b>Angola</b>	58	34	393,745	2012, 2013, 2019
<b>Mozambique</b>	70	42	380,958	2010, 2020
<b>São Tomé and Príncipe</b>	77	32	322,999	2008, 2011, 2012
<b>Total</b>	205	108	1,097,702	

**Table 1:** Profile of the PALMA corpora.

For the data set used in this paper, we extracted all the occurrences of *ir* ‘to go’ and *chegar* ‘to arrive’ from the three corpora on the CQPweb platform and then proceeded to exclude the following contexts:

- (i) contexts without a locative argument.
- (ii) periphrastic verbal constructions of the type *ir/chegar* (constituent) + main verb.
- (iii) preposed locative arguments (antecedents of relative clauses, topics, focused locatives, etc.), which are more prone to preposition deletion.
- (iv) prepositioned locative arguments that were separated from the verb by locative adverbs, such as *lá* ‘there’ and *ali* ‘there’.
- (v) lexicalized constructions involving *ir* and *chegar*.
- (vi) unclear contexts, for example due to hesitations, reformulations, etc.
- (vii) contexts produced by two interviewers who are native speakers of AP and MP.

The remaining constructions were annotated with respect to (i) the preposition selecting the locative argument (V PP<sub>Loc</sub>) and cases without preposition (V NP<sub>Loc</sub>), (ii) the physical or non-physical nature of the propositions described by predications, and (iii) the nature of the NP (area or container) of PPs headed by locative *em*.

### 3. Goal arguments in AVPs

The available literature on AP, MP, and STP refers the use of a non-standard prepositional strategy with locative *em* ‘in’ to introduce arguments with the thematic role of Goal<sup>1</sup> (e.g., Adriano, 2014, pp. 334–340; Cabral, 2005; Chavagne, 2005, pp. 223–227; Mingas, 2000, pp. 75–77; P. Gonçalves & Chimbutane, 2004; P. Gonçalves, 2010; R. Gonçalves, 2010) instead of the standard prepositions *a* and *para*. Chavagne (2005, p. 227) mentions that the use of *em* with verbs of movement represents a strong tendency, comparing it with Brazilian Portuguese (BP). Based on a smaller spoken corpus, R. Gonçalves (2010, p. 48) shows that in STP there are two main tendencies: *ir* preferentially selects PPs headed by *para* (even in contexts where EP selects *a*), whereas *chegar* shows a tendency toward *em*. Our data show Goal arguments of *ir* and *chegar* (among other verbs) headed by *em* are indeed attested in all three AVPs, as illustrated in (1) below.<sup>2</sup>

---

<sup>1</sup> Other uses of *em*, especially in AP, are also mentioned by these authors, for example, to introduce Sources and Recipients. In a preliminary survey on Goal and Source arguments of verbs of movement based on a subpart of the same corpus, Hagemeyer et al. (2019) noted that Source arguments are quite consistently introduced by preposition *de* in all three AVPs.

<sup>2</sup> *Em* contracts with masculine and feminine definite articles (and with other determinants and pronouns), yielding, respectively *no* and *na* (or its plural counterparts *nos* and *nas*).

- (1) (a) *Logo que cheguei **no** aeroporto...* (AP, corpus data)  
 soon that I.arrived in.the airport  
 ‘As soon as I arrived at the airport...’
- (b) *Por isso, muitos jogadores negros vão **em** Europa.* (MP, corpus data)  
 therefore many players black go in Europe  
 ‘Therefore, many black players go to Europe.’
- (c) *Eles já aconselham que é para ir **no** hospital.*  
 they already advice that it.is to go in.the hospital  
 ‘Now they advise to go to the hospital.’ (STP, corpus data)

In our corpora of the three AVPs, however, Goals introduced by prepositions *a* or *para*, the standard strategies in EP, overall represent the most common strategy, which can be seen in examples (2–3).

- (2) (a) *Quando uma pessoa chega **à** escola...* (AP, corpus data)  
 when a person arrives to.the school  
 ‘When a person arrives at school...’
- (b) *conseguiu ir **à** África do Sul...* (MP, corpus data)  
 he.managed go to.the South Africa  
 ‘he managed to go to South Africa...’
- (c) *Já fui **a** Santa Catarina.* (STP, corpus data)  
 already I.went to Santa Catarina  
 ‘I’ve been to Santa Catarina.’
- (3) (a) *A filha dele foi **para** universidade.* (AP, corpus data)  
 the daughter of.him went to university  
 ‘His daughter went to university.’
- (b) *Amanhã de manhã vou logo **para a** escola.* (MP, corpus data)  
 tomorrow of morning I.go right away to the school  
 ‘Tomorrow morning I’m going to school right away.’
- (c) *Fui **para** um outro meio em que tinha cabo-verdianos.*  
 I.went to a other place in that had Cabo Verdeans  
 ‘I went to another place where there were only Cabo Verdeans.’  
 (STP, corpus data)

The overall results for the three AVPs with respect to the use of *a*, *para*, *em*, *até*,  $\emptyset$  (null preposition), and other prepositioned contexts introducing Goal complements of the verb *ir* in our corpora are presented as percentages and the corresponding raw numbers (between parentheses) in **Table 2** below.

<i>Ir</i>	<b>PPa</b>	<b>PPpara</b>	<b>PPem</b>	<b>PPaté</b>	<b>Ø (NP)</b>	<b>Other</b>	<b>total</b>
<b>AP</b>	27,90% (60)	38,14% (82)	29,30% (63)	2,33% (5)	1,86% (4)	0,47% (1)	100% 215
<b>MP</b>	51,53% (252)	32,32% (158)	9% (44)	3,47% (17)	1,02% (5)	2,66% (13)	100% (489)
<b>STP</b>	30,64% (129)	61,52% (259)	1,90% (8)	4,28% (18)	0,47% (2)	1,19% (5)	100% (421)
<b>Total</b>	39,2% (441)	44,36% (499)	10,22% (115)	3,55% (40)	0,98% (11)	1,69% (19)	100% (1125)

**Table 2:** Selection of Goal arguments of *ir* ‘to go’ in three AVPs.

The overall results for the three AVPs with respect to the use of *a*, *em*, *até*,  $\emptyset$ , and other prepositioned contexts introducing Goal complements of the verb *chegar* are shown in **Table 3** below.

<i>Chegar</i>	<b>PPa</b>	<b>PPem</b>	<b>PPaté</b>	<b>Ø (NP)</b>	<b>Other</b>	<b>total</b>
<b>AP</b>	43,4% (23)	39,62% (21)	7,55% (4)	1,88% (1)	7,55% (4)	100% (53)
<b>MP</b>	25,79% (33)	67,19% (86)	2,34% (3)	2,34% (3)	2,34% (3)	100% (128)
<b>STP</b>	37% (37)	37% (37)	4% (4)	22% (22)	0% (0)	100% (100)
<b>Total</b>	33,10% (93)	51,25% (144)	3,91% (11)	9,25% (26)	2,49% (7)	100% (281)

**Table 3:** Selection of Goal arguments of *chegar* ‘to arrive’ in three AVPs.

The tables show in the first place that the varieties exhibit substantial disparity with respect to the patterns selected by *ir* and *chegar*, both within and across varieties. It also follows that the complements of *ir* and *chegar* are predominantly PPs and not NPs. STP, however, is exceptional in the sense that this variety exhibits a rather expressive number of cases where *chegar* selects an NP argument (22/22%).

In the specific case of *ir* (cf. **Table 2**), the standard EP strategies with *a* and *para* are most frequent (940/83,56%), with a strong preference for *para* in the case of STP (259/61,52%) in

comparison to AP and MP.<sup>3</sup> With respect to the use of the non-standard strategy with *em*, which only represents a relatively small number of overall cases in these varieties (115/10,22%), AP is the variety that stands out for its more widespread use of this preposition (63/29,30%). The application of Fisher's test<sup>4</sup> (Eddington, 2016) to the data in **Table 2** shows that significant differences can be observed between varieties, since the p-value is much lower than 0.05 in all cases: The pairs AP/MP, AP/STP, and MP/STP all have a  $p < .001$ .

In the case of *chegar* (cf. **Table 3**), the variation in the AVPs mainly occurs between two prepositional strategies, either with *a* or with *em*. In contrast to *ir*, the non-standard strategy with *em*, corresponding to 144/51,25% of the occurrences, is more widespread than the canonical strategy, which accounts for 93/33,10% of the cases. While the strategy with *em* is dominant in MP (86/67,19%), it is also common in AP (21/39,62%) and STP (37/37%). Differently from AP and MP, as mentioned above, STP also shows a fair number of cases where *chegar* selects an NP (22/22%), a tendency that was not observed with regard to *ir* (2/0,47%). Applying Fisher's test to the data in **Table 3** shows, again, that significant differences between varieties can be observed: the pair AP/MP has a p-value = 0.006181; the pair AP/STP a  $p = 0.004861$ ; and, most notably, the pair MP/STP a  $p < .001$ .

All in all, the statistical tests suggest significant differences between the three varieties with respect to the selection of Goal complements with both *ir* and *chegar*.

#### 4. Location in the main contact languages

In order to assess the role of language contact, we will now turn to the strategies used to express Goal arguments in the main contact languages of the three urban AVPs at stake, which are the Bantu languages Kimbundu (Mbundu) and Changana, for AP and MP respectively, and the creole language Forro for STP. If contact plays a role, as has been claimed by several authors, the role of these languages is expected to be both direct and indirect. Direct, in the sense that there are speech communities and/or bilinguals of these languages within the urban environment and among our informants, especially in the case of MP; indirect, in the sense that these languages – or sometimes languages with a similar typology spoken by migrants moving to urban areas – have been historically in contact with Portuguese, even if speakers do not necessarily speak them (fluently or actively).

---

<sup>3</sup> The use and distribution of PPs headed by *a* and *para* when selected by *ir* (and other verbs of directed motion) in these varieties is a topic for future research.

<sup>4</sup> The Fisher's exact test is a statistical test used to determine whether nonrandom associations hold between two categorical variables, for example between AP and STP. A decision is made based upon the p-value threshold (usually .05). A  $p < .05$  means no association, i.e., a significant difference is detected.

For Proto-Bantu, the ancient ancestor of the contemporary Bantu languages, three locative noun class prefixes have been reconstructed, corresponding to classes 16, 17, 18 (\**pà-*, \**kù-*, \**mù-*), which have a reflex in many contemporary Bantu languages but have been lost or only occur as vestiges in other ones (e.g., Zeller, in press).

Kimbundu, a western Bantu language, preserves reflexes of the three proto-Bantu locative noun classes (e.g., Araújo, 2013; Chatelain, 1888–1889; Diarra, 1990; Miguel, 2019).<sup>5</sup> The marker that is standardly used to introduce the Goal argument of verbs of directed motion is *ku*, which stems from class 17, as illustrated in (4–5).

(4) *Ngiya ku {ngeleja/kalunga/Lisboa}*. (Kimbundu, Afonso Miguel, p.c.)  
 1SG.go.FV LOC {church/sea/Lisbon}  
 ‘I’m going to {the church/the sea/Lisbon}.’

(5) *Nzwa watula {ku Wambu / ku ngeleja}*.  
 Nzwa 3SG.PST.arrive.FV LOC Huambo / LOC church  
 ‘Nzwa arrived in Huambo / at the church.’ (Kimbundu, Afonso Miguel, p.c.)

Differently from Proto-Bantu, however, the locative markers in Kimbundu exhibit properties of free morphemes, i.e., prepositional items, instead of prefixes, as shown in (4–5). The prepositional use of locative morphemes has been demonstrated for a number of Bantu languages in different areas (e.g., Marten, 2010; Zeller, in press). Although the prepositional status of locative markers in Kimbundu has been assumed in different studies without being substantiated with linguistic evidence (e.g., Diarra, 1990, p. 59), it is, for instance, supported by the fact that the locative marker does not trigger locative agreement in object position. In (6), it is the noun *ngeleja* ‘church’ which triggers class 9 agreement with demonstratives and possessives and not the locative related to class 17.

(6) *Ngiya ku {ngeleja yiyi / yami}*. (Kimbundu, Afonso Miguel, p.c.)  
 1SG.go.FV LOC(17) 9.church 9.DEM / 9.POSS  
 ‘I’m going to {this/my} church.’

Furthermore, when the locative marker selects an animate object, a connective morpheme *-a* is required, as illustrated in (7).<sup>6</sup> This feature is not compatible with a prefixal status of *ku* because the intervention of this type of morphological material between a noun class prefix and the nominal root is not attested in Bantu languages.

<sup>5</sup> Other well-known Bantu languages in Angola whose speakers have migrated to the capital Luanda, such as Umbundu and Kikongo, also exhibit three similar locative classes.

<sup>6</sup> In fact, Kimbundu exhibits differential object marking, which further follows from the fact that locative marker *ku* also introduces arguments with the semantic role of Source and Recipient, requiring the presence of morpheme *-a* when objects are [+ANIM].



- (7) *Ngiya {ku kalunga / kua Phetele}*. (Kimbundu, Afonso Miguel, p.c.)  
 1SG.go.FV LOC sea / LOC.CONN Phetele  
 ‘I went to the sea / to Phetele’s (place).’

The expression of location in Changana, a southeastern Bantu language, is quite distinct from the facts observed for Kimbundu, since the language only exhibits remnants of the three locative classes reconstructed for Proto-Bantu (Chimbutane, 2002; Ngunga & Simbine, 2012; Siteo, 2001). Instead, Goals are characterized by other morphological strategies (Chimbutane, 2002; Siteo, 2001), typically the attachment of suffix *-ini* (or its allomorph *-eni*) to the NP, accompanied by optional prefixation of *á-* or *é-*.

- (8) *Bilá átáya áxikólwéni múndzuku...* (Changana, Siteo, 2001, p. 93)  
 Bila 1.FUT.go LOC.school.LOC tomorrow  
 ‘Bila will go to school tomorrow...’

Siteo (2001, pp. 6–7) mentions that Changana locatives are generally vague with respect to their localization. He illustrates this claim with the example *nambzeni* (*nambu* ‘river’ + *-eni*), which “can refer to a location inside the river, on the river, at the bank of the river, or in the general area of the river.”

Changana further displays a number of inherently locative nouns, which includes borrowings, that lack any overt marking, as *mánánga* ‘desert’ in (9a), or can be prefixed with *á-* to reinforce their locativity, as *ákáyá* ‘home’ in (9b).

- (9) (a) *Makamela máyé mánánga* (Changana, Siteo, 2001, p. 9)  
 6.camels 6.go.PT desert  
 ‘The camels went to the desert.’  
 (b) *Podíná ámúka ákáyá.* (Changana, Siteo, 2001, p. 9)  
 Podina 1.go:back.PR LOC.home  
 ‘Podina is going back home.’  
 (c) *Bilá áfikélé káyá.* (Changana, F. Chimbutane, p.c.)  
 Bila 1.arrive.APPL.PT home  
 ‘Bila arrived home.’

For endogenous toponyms, on the other hand, prefix *ká-* is employed (10a), whereas toponyms of foreign origin are not marked or are optionally marked by prefix *á-* (Siteo, 2001, p. 10) as shown in (10b).

- (10) (a) *Bilá áya káTembe.* (Changana, Siteo, 2001, p. 11)  
 Bila 1.go.PR LOC.Tembe  
 ‘Bila goes to Catembe.’

- (b) *Bilá áya áJoní.* (Changana, Siteo, 2001, p. 11)  
 Bila 1.go.PR LOC.Johannesburg  
 ‘Bila goes to Johannesburg.’
- (c) *Bilá áfikélé áMapútu.* (Changana, Siteo, 2001, p. 135)  
 Bila 1.arrive.APPL.PT LOC.Maputo  
 ‘Bila arrived precisely at Maputo.’

Finally, Goal arguments in Forro, the main creole language of S. Tomé, are uniformly expressed as NPs (e.g., Hagemeyer, 2004).

- (11) (a) *Kuma n ga ba fesa sun alê?* (Forro, Hagemeyer et al., 2014)  
 how 1SG FUT go party Mr. king  
 ‘How will I go to the party of the king?’
- (b) *N ba glêza ku inen.* (Forro, Hagemeyer et al., 2014)  
 1SG go church with 3PL  
 ‘I went to church with them.’
- (c) *Oze n ga ba Txindadji.* (Forro, Hagemeyer et al., 2014)  
 today 1SG FUT go Trindade  
 ‘Today I will go to (the town of) Trindade.’
- (12) (a) *San xiga palaxu.* (Forro, Hagemeyer et al., 2014)  
 3S.FRM arrive palace  
 ‘She arrived at the palace.’
- (b) *Florensa xiga poson.* (Forro, Hagemeyer et al., 2014)  
 Florença arrive city of São Tomé  
 ‘Florença arrived at the city of São Tomé.’
- (c) *Tudu ngê xiga misa.* (Forro, elicited)  
 all people arrive mass  
 ‘Everybody arrived at the mass.’

In sum, the three main contact languages of the urban AVPs analyzed exhibit fundamentally different solutions to encode Goal arguments with verbs of directed motion: Kimbundu exhibits PPs headed by *ku*; Changana uses a strategy in which NPs are morphologically marked; in Forro, Goals are also NPs but lack any (overt) morphological marking.

## 5. Discussion

A cross-comparison between the patterns available in the AVPs and their respective main contact languages shows that the contact-induced hypothesis runs into several problems.

First, the AVPs in this study predominantly use a PP strategy to introduce Goals of the direction motion verbs *ir* and *chegar*, whereas two of the contact languages, Changana and Forro, exhibit an NP strategy, respectively with or without overt morphology. Second, while the prepositions heading PPs in the AVPs show considerable variation, a clear-cut distinction arises between *ir*, where standard *a* and *para* constitute the dominant pattern, and *chegar*, which shows a preference for the non-canonical strategy with *em*. The main contact languages (Kimbundu, Changana, and Forro), on the other hand, use a language-internally uniform strategy to express the complement of these and other directed motion verbs.<sup>7</sup>

The contact-induced hypothesis is also unable to explain why *ir* in STP shows a strong preference for *para* and hardly any cases of *em*, and *chegar* a primary pattern with *em*, since Forro systematically uses an NP argument in both cases. In AP, however, the more pronounced overall presence of *em*, although far from being the exclusive patterns, would be more consistent with a more emphatic role of language contact, since Kimbundu also displays a generalized and uniform PP strategy with locative *ku*, which could further facilitate syntactic transfer and convergence.<sup>8</sup> This line of argumentation can arguably not hold for MP though, since it was shown that its main contact language Changana exhibits morphological strategies.

To overcome this problem in the case of MP, P. Gonçalves and Chimbutane (2004) and P. Gonçalves (2010) propose an analysis, based on a scenario of L1 Bantu speakers who are L2 learners/acquirers of Portuguese, whereby preposition *em* is reanalyzed as an NP-internal locative Case marker, [*em*-NP], i.e., as a prefix, inspired in the Bantu noun class tradition. This hypothesis, they argue, would then be able to account for examples like (13a-b), where preposition *no* in (13a) and *em* (in 13b) would be incorporated into the NP as follows: [<sub>NP</sub> *no-encontro da igreja*] and [<sub>PP</sub> *para* [<sub>NP</sub> *em-casa*]].

- (13) (a) *you no encontro da igreja* (MP, corpus data)  
 I.go in.the meeting of.the church  
 ‘I’m going to the church meeting’
- (b) *leva os meus filhos para em casa do outro homem*  
 you.take the my children to in house of.the other man  
 ‘you take my children to the other man’s place’ (MP, corpus data)

---

<sup>7</sup> Source arguments of directed motion verbs constitute yet another argument against the contact-induced hypothesis, since the contact languages use similar strategies for the selection of Goals and Sources. However, as mentioned in footnote 1, Source arguments in the AVPs are typically headed by preposition *de*, hereby converging with the pattern in European Portuguese.

<sup>8</sup> In addition, the overlap between KB *ku* and AP *em* can also be observed in the marking of dative objects (Recipients), which constitutes an additional argument for a greater role of the contact-induced hypothesis (cf. R. Gonçalves, Duarte & Hagemeyer, 2022).

We concur with P. Gonçalves and Chimbutane (2004) that examples of cooccurring prepositions, in (13b), in particular *para* and *em* are rare. In our data, we identified only 5 such occurrences, produced by 4 different speakers. In 4 of these occurrences, the locative NP was *casa* ‘house’. The scarcity of these constructions would therefore not only weaken the hypothesis that *em* is a prefix, but would also require further explanations as to the workings of the grammaticalization of a preposition into a bound morpheme in MP. As shown in section 4, the primary locative strategy in Changana relies on a suffix (*-ini* or *-eni*), and not on a prefix. Moreover, there is no evidence that other Bantu affixes, for example negation prefixes or verbal extensions (applicatives, etc.), induce reanalysis of Portuguese morphemes as bound morphology in MP (or AP). This raises the more general issue as to whether or how morphology in a L1 is processed and transferred into an L2 target language in contact situations. In the case of Changana-Portuguese it would be a costly operation, requiring both a categorial and syntactic reanalysis of *em*.

Furthermore, Bantu locative morphemes are typically prefixed to the head noun, as in the case of Changana’s *ká-* and *á-* described in section 4, which means that in a strong version of the contact hypothesis the strategies with *em* are not expected to contract with definite articles<sup>9</sup> or demonstratives nor to be separated from the head noun by possessives, numerals, prenominal adjectives, etc., contrary to fact (e.g., *há quem chega no meu salão* ‘some people arrive at my beauty salon’, MP corpus data). This type of noun modifying material is generally postnominal in Bantu, but in those languages where it occurs preverbally, the locative marker is a free morpheme (preposition).<sup>10</sup>

The contact hypothesis for MP was also adopted by Avelar and Álvarez López (2018) for the (L2) Cabinda variety of AP, which is primarily in contact with Ibinda, a western Bantu language belonging to the Kongo cluster, which exhibits prenominal locative morphemes. While this would be syntactically consistent with contact-induced hypothesis above, it still faces the problems mentioned in the previous paragraph regarding the NP-internal syntactic differences between Bantu and Portuguese.

Finally, the NP-internal prefixation hypothesis proposed originally by P. Gonçalves and Chimbutane (2004) is unable to account for AP, since it was shown that Kimbundu locative markers introducing Goals are free morphemes (prepositions). Since Kimbundu and Portuguese both exhibit prepositions, it would be counterintuitive to claim that Ps incorporate in AP with verbs of directed motion. In fact, as expected, we did not find any instance of double prepositions in our AP data.<sup>11</sup>

---

<sup>9</sup> Our MP corpus displays no occurrences of *para* + *no(s)/na(s)*.

<sup>10</sup> See, for instance, Chimbutane (2002, pp. 135–142) on the distinction in Changana between locative prefix *ká-* and free morpheme *ka*, which exhibit different morphological, syntactic, and semantic properties.

<sup>11</sup> Avelar (2017, p. 27) also mentions that no instances of double prepositions were found in Cabinda Portuguese.

The gradual withdrawal of the contact languages caused by shift, especially in the urban environment in Angola and São Tomé, where Portuguese is not only widely spoken but also frequently the native or primary language of the speakers, with increasing monolingualism in Portuguese, especially among the younger generations, constitutes yet another argument that increasingly weakens the role of language contact with respect to the observed patterns. This tendency has become stronger since the independences in 1975 and can be correlated with the massification of education in the official language and, concomitantly, higher educational attainment. Even though the spoken local (non-native) varieties constitute the main input (e.g., Stroud, 1997), the role of education is expected to raise the speakers' linguistic awareness and to promote wider use of patterns found in the standard language.<sup>12</sup> While the historical relevance of language contact cannot be ignored, the contemporary urban environments have become the playground of thriving language shift toward Portuguese, diminishing the role of transfer of substrate patterns.

Finally, compared to EP, the AVPs appear to privilege prepositions *para* and *em* over *a* with the verbs at stake, even though ultimately a finer-grained analysis is required to assess the use of *para* and *a*, for example with regard to implications concerning a longer or a shorter stay at the Goal, as in the difference between *ir para Luanda* 'to go to Luanda' vs. *ir a Luanda* 'to go to Luanda', where the former implies a longer stay in Luanda than the latter. The partial loss of *a* is an expected outcome in contact situations where L2 acquisition plays or played a prominent role. Phonetically weak and semantically opaque functional items, such as directional preposition *a*, definite articles, or accusative clitics, are prone to loss and/or partial replacement by other, often less functional material. In contexts of ambiguous and insufficiently dense input (e.g., P. Gonçalves, 2002, 2004), it is expected that directional *para* or locative *em* have the upper hand in comparison to *a*. For instance, the latter, unlike the former two, did not make it into the Portuguese-related creoles in West-Africa belonging to the Upper Guinea and Gulf of Guinea clusters. Moreover, in (spoken) BP, which also has a history of substantial L2 acquisition and partial restructuring (e.g., Lucchesi, 2001), directional preposition *a* has given rise to variation, favoring in particular *em* in the structures at stake, and does not occur in Afro-Brazilian subvarieties which underwent substantial restructuring (see Avelar, 2017, for discussion).

In sum, the empirical evidence does not underscore a major *direct* role for language contact regarding the selection properties of directed motion verbs *ir* and *chegar*, even though there may be – on a limited scale – individual grammars that privilege the grammatical properties of the contact languages. It was also shown that a morphological analysis based on language

---

<sup>12</sup> In a study carried out by Firmino (2002, *apud* P. Gonçalves, 2005), it is mentioned that Mozambicans have positive attitudes toward the EP grammar, except toward its pronunciation, which is considered snob.

contact, proposed initially for MP, cannot be applied across the board, primarily because it is not consistent with the data from the contact languages and the AVPs.

In light of the above, we aim to develop a different, semantic-oriented approach to the data, as an alternative (and complement) to the contact-induced hypothesis.

## 6. A semantic hypothesis

In the remainder of this paper, we discuss Goal PPs introduced by *em*, a non-standard strategy in EP, focusing primarily on the lexical semantics of verbs and their objects and on the semantics of predications.

### 6.1. Goals expressed by locative phrases

Goals are not always expressed by Goal prepositions, such as *a* and *para* in EP. In English, for instance, locative phrases can be understood as goals and this interpretation seems to arise irrespective of the semantics of the verb, that is, irrespective of being a directed motion verb or a manner of motion verb.

- (14) (a) John walked in the room.  
 (b) Kim jumped on the bed. (Beavers, Levin & Tham, 2010, p. 33)

Although both *in* and *on* in (14) are locative prepositions, besides the locative interpretation, the PPs they introduce also have directional (Goal) interpretations, which are easier to achieve in the right context, namely if the entity referred to is standing outside the room, in (14a), or next to the bed, in (14b). Levin, Beavers and Tham (2009) argue that this strategy of combining motion verbs with locative PPs is one of the ways of expressing directionality with motion verbs in English – the other one is combining motion verbs with Goal-marking PPs – but there are differences in their productivity, as the former strategy shows acceptability variation whereas the latter is widespread.

This possibility of Goal readings of locative phrases is also available in other languages, such as Afrikaans, Ancient Greek, Dutch, Norwegian, Russian, and Ukrainian (cf., e.g., Beavers, Levin & Tham, 2010; Nikitina & Maslov, 2013), as well as Romance languages, as French, Spanish, or Italian (e.g., Nikitina, 2008; Ursini, 2013).

- (15) Rick sprong in het meer. (Dutch, Gehrke, 2007, p. 248)  
 Rick jumped in the lake (locative / directional)  
 ‘Rick jumped in the lake.’
- (16) Max a couru dans sa chambre. (French, Nikitina, 2008, p. 177)  
 ‘Max ran into his room.’

In fact, Nikitina (2008, 2017) points out that in many languages, including Romance languages, (static) Locations and Goals tend to be expressed by the same linguistic forms (e.g., prepositions, case), being the distinctions between both readings context dependent. Also, Pantcheva (2010) surveys typologically distinct languages and concludes that, despite the general tendency toward a distinct marking of Locations, Goals and Sources (60% of the languages under scrutiny use differential marking), there is also a considerable number of languages (approximately 30%) exhibiting the same strategies for the expression of both Locations and Goals.

If we turn to the two main Portuguese varieties, the situation is twofold. On one hand, in BP, the locative preposition *em* ('in', 'on') is used with directional interpretation in events of movement that typically include verbs of directed motion.<sup>13</sup>

- (17) (a) {*Fui/cheguei/vim*} *no cinema*. (°BP; \*EP) (Avelar, 2017, p. 16)  
 I {went/arrived/came} in.the cinema
- (b) *Fui/cheguei/vim ao cinema*. (°BP; °EP) (Avelar, 2017, p. 16)  
 I {went/arrived/came} to.the cinema  
 Both: 'I {went to/arrived at/came to} the cinema.'

In EP, on the other hand, only a small number of verbs denoting some sort of change of location can appear with preposition *em* introducing Goal complements: this short list includes *entrar* ('to enter') (cf. (18a)), *aparecer* ('to appear'), *esconder* ('to hide'), *penetrar* ('to penetrate') and *pôr* ('to put') (cf., e.g., Raposo & Xavier, 2013, pp. 1543–1544). The preposition *em* conveying a directional meaning cannot be used with other directed motion verbs (nor manner of motion verbs), as illustrated in (18b-c).

- (18) (a) *Entrei em casa*. (EP)  
 I.entered in house  
 'I entered the house.'
- (b) *Subi no quarto*. (EP)  
 I.went.up in.the room  
 'I went up in the room.' / \*'I went up to the room.'
- (c) *Caminhei em casa*. (EP)  
 I.walked in house  
 'I walked inside the house.' /\*'I walked to the house.'

---

<sup>13</sup> Although this clashes with the BP standard (cf., e.g., Avelar, 2017; Farias, 2006; Vieira, 2009).

All in all, the occurrence of locative prepositions with a Goal meaning is crosslinguistically widespread, EP being an exception.<sup>14</sup> In the next section, we review some hypotheses for this directional reading of locative prepositions.

## 6.2. Hypotheses for locative phrases with directional meaning

### 6.2.1. Brazilian Portuguese preposition *em*

There are a few sociolinguistic studies on BP geographic varieties that try to explain the directional meaning of *em* in terms of pragmatic and/or semantic factors (e.g., Mollica, 1996; Vieira, 2009; Wiedemer, 2008). These studies scrutinize the factors underlying the use of *em* in directional contexts with certain verbs of directed motion and converge in postulating that directional *em* is favored by the fact that preposition *em* with a directional interpretation typically introduces a closed space, like a *house*, whereas directional prepositions *a* ‘to’ and *até* ‘up to’ typically introduce open spaces, such as a *beach*.

More recently, Rammé (2017) and Ferreira and Basso (2019) discuss the use of directional *em* in BP using a nanosyntactic framework. In both proposals, the authors assume that prepositions can be either locative or directional and, when a locative preposition has a directional interpretation, this interpretation depends on the verb (that is, there is a structural effect). Using this point of departure, they postulate that *em* is always a locative preposition, and that the PP only has a locative meaning, i.e., *em* only codifies location, not direction, even in those cases where a directional interpretation arises. The directional interpretations of *em* depend exclusively on the verb’s meaning, which explains why directional *em* only occurs with verbs that codify directed motion. Therefore, both studies conclude that the directional interpretation of BP *em* is a consequence of structural ambiguity, or false syncretism (cf. Gehrke, 2008). Additionally, Rammé (2017) also posits that *em* in BP has not changed its semantic value in the past two centuries (contrary to what is usually assumed in the literature), whereas Ferreira and Basso (2019) propose that *em* is used in those contexts where the speaker wants to assure that, at the end of the event, the Figure is in the interior of the Goal (cf. the abovementioned pragmatic studies of BP *em*).

---

<sup>14</sup> The directional use of *em* is not a feature of EP. In addition to the fact that this use is not mentioned in reference work (grammars and dictionaries) on EP, it is not found in corpora, such as the Reference Corpus of Contemporary Portuguese (CRPC) or the CETEMPúblico newspaper corpus, and native speakers consider it ungrammatical. On the other hand, EP apparently exhibits the opposite phenomenon, that is, locative uses of the directional preposition *para*, as in *O João está para o Algarve* (‘John is (somewhere) in Algarve’). See Oliveira et al. (2021) for a description of the locative readings of preposition *para* with different copula verbs (*ser*, *estar*, and *ficar*).



### 6.2.2. English preposition *in*

Although there are proposals explaining the occurrence of locative PPs with directional meaning that follow a lexical approach (e.g., Alonge, 1997; Fábregas, 2007; Folli & Ramchand, 2005), Levin, Beavers and Tham (2009) and Beavers, Levin and Tham (2010) convincingly argue that the occurrence of these locative PPs with directional interpretation can be explained taking into consideration semantic and pragmatic factors, following Nikitina (2008). In general, the research on these matters points out two strategies for the encoding of the directional meaning: (i) to encode the directional meaning in a specialized satellite, if available; this kind of Goal combines with all types of verbs, including manner verbs; (ii) to encode the directional meaning in a specialized class of inherently directional verbs (typically manner verbs are excluded). However, Nikitina (2008) suggests a third option to describe directed motion. In fact, languages lacking specialized means of encoding directionality outside the verb (by specialized satellites) do not always restrict the expression of directed motion to a subset of inherently directional verbs; instead, the directional meaning is not expressed overtly, that is, the strategy of expressing directed motion relies on contextual inference rather than lexical encoding (“zero” encoding strategy; cf. Kopecka, 2009, p. 55 for a similar claim).

Nikitina’s proposal relies on a compositional view of directional meaning. Thus, she assumes that more than one element in the sentence can contribute to the directional interpretation and that the directional meaning does not have to be encoded by specialized lexical material to be inferred. Furthermore, the choice of a strategy partly depends on factors that have to do with alternative ways of conceptualizing motion events. One of those factors is context. As Beavers, Levin and Tham (2010, p. 33) posit, “... locative phrases are understood as goals precisely in those contexts that allow a reader or hearer to infer that a goal interpretation is intended.” In (19), repeated from (14a), “in the room” is interpreted as a Goal phrase only if John is standing nearby the door of the room; otherwise, the phrase is interpreted as locative (the place where the event of John walking occurred).

(19) John walked in the room.

The verb’s lexical semantics is also an important factor in Nikitina’s proposal. Locative PPs with Goal interpretation are found more often with verbs that project non-durative predications; on the contrary, they do not arise alongside durative verbs; therefore, they are usually combined with directed motion verbs (usually punctual, transitional verbs), but not with manner of motion verbs (usually durative verbs). Similar conclusions grounded on English, Dutch, and French data can be found in Gehrke (2007),<sup>15</sup> Thomas (2004), and Kopecka (2009).

---

<sup>15</sup> “The ambiguity of sentences with *in* and *on* between a locative and a directional reading is not observed with all instances of these Ps, though. In particular, only certain verbs such as *kick*, non-iterative *jump*, *throw*, *put*, *fall*, among others, henceforth kick-verbs, can trigger a directional reading (...). With other motion verbs like *dance*, *crawl*, *walk*, *swim*, among others, henceforth swim-verbs, these prepositions only get a locative reading.” (Gehrke, 2007, p. 247)

The last relevant factor underlying the expression of Goals in Nikitina (2008) is the semantics of the complement of the preposition. The author argues that locations with well-defined boundaries (“containers”<sup>16</sup>), that is, that are not surrounded by a transitional zone,<sup>17</sup> are used more often in locative PP with directional interpretation; on the contrary, the locations that do not have well-defined boundaries (“areas”<sup>18</sup>) are excluded from these constructions (cf. also Kopecka, 2009, for French).

Although Nikitina’s proposal concerns English data, there is evidence that the semantic/pragmatic account of locative Goals can be extended to Romance languages. For instance, Kopecka (2009) analyses some manner of motion verbs with locative prepositions in French and concludes that the factors that favor the directional readings of locative prepositions in English can be extended to account for French data. In what concerns BP *em*, the studies of Mollica (1996), Wiedemer (2008), and Vieira (2009) also adopt the same analysis, emphasizing the influence of the semantics of the complement of the preposition, while Rammé (2017) and Ferreira and Basso (2019) highlight the importance of the meaning of the verb. In the next section it will be shown that certain tendencies observed with respect to directional readings of *em* in the AVPs can be accounted for by Nikitina’s hypothesis.

### 6.3. A semantic approach to directed motion in AVPs

#### 6.3.1. Introducing a semantic feature hypothesis

Following Nikitina’s (2008) semantic/pragmatic approach of directional locatives, we argue that the directional readings of preposition *em* in AVPs are favored:

1. when the speaker wants to describe an event of movement, that is, an event in which there is an entity (the Figure) that moves toward a reference object (the Ground), and this movement can be described by a series of spatial coordinates.
2. when the reference object has well-defined boundaries.
3. when the predication describes a non-durative change of place.

Concerning 1., one of the key aspects of the semantic/pragmatic account of the directional readings of locative phrases is that it involves actual movement in the “real world”. In other words, locative phrases are interpreted as Goal phrases if, during the event, the Figure changes its location and, in the end of the event, the Figure is located typically inside that location. Therefore, one assumes that the Figure must be located by spatial coordinates. However, verbs of

---

<sup>16</sup> Some examples: *rooms, buildings, boxes, cars, water, ground*.

<sup>17</sup> A transitional zone is “perceived as neither contained by the location nor located outside of it.” (Nikitina, 2008, p. 186)

<sup>18</sup> Some examples: *cities, mountains, countries, space, forests, neighborhoods*.

movement can also be used to describe situations that are events that do not correspond to actual movement in the real world or do not involve movement at all, as in (20) and (21).

(20) The price of gasoline rose from \$2.6 to \$3.

(21) The road goes from the village to the mountain.

In (20), there is an event, and, during this event, there is a change in a scale (associated with the word “price”), and not a change in the physical location of the entity denoted by “the price of gasoline”. As for (21), the predication corresponds to a state describing the (static) location of the entity denoted by “the road”. In the cognitive literature, sentences like (20) are typically treated as metaphors, whereas sentences like (21), that correspond to a description of a static situation, are treated as “fictive motion” (Talmy, 2000), “subjective motion” (Langacker, 1986; Matsumoto, 1996) or “non-actual motion” (Blomberg & Zlatev, 2014).

This difference between events that correspond to a change in the physical location of the Figure (as the entity denoted by “the man” in “the man went to the village”), and, on the other side, events that do not correspond to a change in the physical location, as in (20), or do not correspond to a change at all, as in (21), seems to be relevant to the analysis of predications with verbs of movement. Take, for, instance, the verb *ir*. This verb can cooccur with three Goal prepositions, *a*, *para* and *até (a)*,<sup>19</sup> and one can change one preposition for another if the predication involves physical motion, as in (22), although with subtle changes in the meaning of the sentence. However, if the predication describes a non-physical motion, these prepositions are not interchangeable. See (23), where *ir para o FC Porto* means ‘to start working at FC Porto’ (cf. Leal, Oliveira & Silvano, 2017).

(22) *O rapaz foi {a / para / até} casa.* (EP)  
 the boy went {to / to / up.to} house  
 ‘The boy went home.’

---

<sup>19</sup> Preposition *até* alternates with prepositional locution *até a*, formed by the combination of *até* and preposition *a*, according to the nature of the complement: *até a* is used whenever the NP complement bears a definite article; otherwise, simple preposition *até* is used (e.g., Raposo & Xavier, 2013, pp. 1504–1505, 1556). Examples (i) and (ii) below show this difference. The noun *Lisboa* is used without article and therefore only *até* can occur. The noun *Porto*, however, requires the use of the masculine, definite article and, concomitantly, the use of *até a*.

(i) *Fui {até Lisboa / \*até a Lisboa}.*  
 I.went up.to Lisbon / up.to to Lisbon  
 ‘I went to Lisbon.’

(ii) *Fui {\*até o Porto / até ao Porto}.*  
 I.went up.to the Porto / up.to to.the Porto  
 ‘I went to Porto (Oporto).’

- (23) *O rapaz foi {\*ao / para o / \*até ao} FC Porto.* (EP)  
 the boy went {to.the / to the / up.to to.the} FC Porto  
 ‘The boy joined FC Porto.’

Therefore, in the analysis of the data of AVPs, we took into consideration this difference between (i) predications expressing events that denote physical motion (cf. (24)) and (ii) predications that express events that do not denote physical motion (cf. (25)) or that express locations (that is, fictive motion).

- (24) *fui à província de Lubango* (AP, corpus data)  
 I.went to.the province of Lubango  
 ‘I went to Lubango province’

- (25) *fui à tropa* (AP, corpus data)  
 I.went to.the army  
 ‘I joined the army’

Concerning 2., that is, the semantics of the complement of the preposition *em*, we consider, as in Nikitina (2008), that the directional readings of preposition *em* are favored when the reference object corresponds to a region with well-defined boundaries, and when there is no transition zone between the region that corresponds to the reference object and all the regions that do not correspond to the reference object. In other words, the directional readings of preposition *em* will be favored if the preposition introduces NPs denoting containers, not areas. We therefore distinguished three kinds of NP complements:

- (i) *areas*: locations that do not have well-defined boundaries, that are surrounded by a transitional zone (some examples from the corpora: *lavra* ‘farmland’, *município* ‘municipality’, *praia* ‘beach’, ...).
- (ii) *containers*: locations with well-defined boundaries, that is, which are not surrounded by a transitional zone (some examples from the corpora: *casa* ‘house’, *bloco* ‘operating room’, *loja* ‘store’, ...).
- (iii) *undefined NP*: this third category is used in those cases that are not adequate for any of Nikitina’s classes (some examples from the corpora: *porta* ‘door’, *árvore* ‘tree’, ...), or if one cannot classify the relevant NPs, for instance, due to lack of context (some examples from the corpora: *sítio* ‘place’, *médico* ‘doctor’, *curandeiro* ‘traditional healer’, ...).

Concerning 3., in Nikitina’s proposal, the lexical information exhibited by the verb form is a relevant factor, because locative PPs with Goal interpretation are found more often with verbs that project non-durative predications (culminations, in Moens’ (1987) terminology) than with durative predications (processes or culminated processes); on the contrary, they should not arise

alongside durative verbs. We therefore tested the *Aktionsart* of the verbs *chegar* and *ir*, namely their durative vs. non-durative nature, using the tests usually found in the literature (cf. e.g., Dowty, 1979; Vendler, 1957; for EP, Leal, 2009), and concluded that *chegar* is a non-durative verb, whereas *ir* is ambiguous between durative and non-durative readings (see Leal, Oliveira & Silvano, 2018).

The following examples, based on the judgement of EP speakers, illustrate these conclusions. Example (26) shows that, with the adverbial “in x time”, the sentence with *ir* entails the truth of the same sentence in the progressive during the same time, which is a diagnostic for durativity. The same test applied to *chegar*, in (27), shows that the predication is non-durative.

- (26) *O rapaz foi para a escola em 10 minutos.*  
 the boy went to the school in 10 minutes  
 ‘The boy went to school in 10 minutes.’  
 → *O rapaz esteve a ir para a escola durante esses 10 minutos.*  
 ‘The boy had been going to school for 10 minutes.’

- (27) *O rapaz chegou à escola em 10 minutos.*  
 the boy arrived at.the school in 10 minutes  
 ‘The boy arrived at school in 10 minutes.’  
 → \**O rapaz esteve a chegar à escola durante esses 10 minutos.*  
 ‘The boy had been arriving at school for 10 minutes.’

Another test for durativity is the occurrence of predications with “at x time” adverbials. In (28), the adverbial *às 14 horas* ‘at two o’clock’ only locates the beginning of the event, which indicates that this event is durative. However, in (19), with *chegar*, the adverbial locates the whole event, which indicates that it is non-durative.

- (28) *O rapaz foi para a escola às 14 horas.*  
 the boy went to the school at.the 14 hours  
 ‘The boy went to school at 2 pm.’
- (29) *O rapaz chegou à escola às 14 horas.*  
 the boy arrived at.the school at.the 14 hours  
 ‘The boy arrived at school at 2 pm.’

However, when combining predications with *ir* with “for x time” adverbials, apparently contradictory results arise. In (30), if the predication is true, then two distinct entailments arise: the boy was going to school during those ten minutes (cf. 30a) or the boy was at school for 10 minutes (cf. 30b). In other words, the adverbial “for x time” can measure the preparatory process (cf. Moens, 1987) of the event (30a) or the consequent state of the event (30b). In the former

case, the entailment is like those that arise with typical durative events (cf. (31)); in the latter case, the entailment is like those that arise with non-durative events (cf. (32)). One can thus conclude that predications with *ir* can have both durative and non-durative readings (see also Leal, Oliveira & Silvano, 2018).

- (30) *O rapaz foi para a escola durante 10 minutos.*  
 the boy went to the school for 10 minutes  
 (a) → *O rapaz esteve a ir para a escola durante 10 minutos.*  
 ‘The boy had been going to school for 10 minutes.’  
 (b) → *O rapaz esteve na escola durante 10 minutos.*  
 ‘The boy was at school for 10 minutes.’
- (31) *O rapaz leu o livro durante 5 minutos.*  
 the boy read the book for 5 minutes  
 → *O rapaz esteve a ler o livro durante 5 minutos.*  
 ‘The boy was reading the book for 5 minutes.’ (Leal, Oliveira & Silvano, 2018, p. 341)
- (32) *O rapaz desmaiou durante 5 minutos.*  
 the boy passed out for 5 minutes  
 → *O rapaz esteve desmaiado durante 5 minutos.*  
 ‘The boy was passed out for 5 minutes.’ (Leal, Oliveira & Silvano, 2018, p. 341)

Before we draw our attention to the analysis of AVP data concerning these parameters, some remarks on Nikitina’s (2008) proposal are in order. Among the relevant parameters favoring the occurrence of directional readings of locative *in*, one also finds co-text. In Nikitina’s study of English, a Goal interpretation of *in* is impossible if there is also a PP indicating the Source of the movement (irrespective of the verb). We therefore checked the occurrences of directional *em* searching for Source phrases. However, there is only one occurrence of a Source phrase in the analyzed data (see (33)), which leads to the conclusion that these verbs (*ir* and *chegar*) do not easily occur with this type of phrases. Notice that, in (33), the Goal phrase is headed by preposition *a*, and not by *em*.

- (33) *da Suazilândia fui à África do Sul* (MP, corpus data)  
 from.the Swaziland I.went to.the South Africa  
 ‘from Swaziland I went to South Africa’

A quick verification of the data available in the PALMA corpora shows that Source phrases can occur with other directed motion verbs and are able to cooccur with Goal phrases headed by *para*, as shown in (34–35).

(34) *os pais vieram de Cabo Verde para cá, né* (STP, corpus data)  
 ‘the parents came from Cabo Verde to here right  
 ‘his parents came from Cabo Verde to here, isn’t it’

(35) *vieram da província para cidade* (AP, corpus data)  
 they.came from.the province to city  
 ‘they came from the province to the city’

So, crucially, what seems to be ruled out is not the cooccurrence of a Source phrase with a Goal phrase, but the cooccurrence of a Source phrase with a Goal phrase headed by *em*, which seems to support Nikitina’s proposal. However, a further analysis based on more data is required to confirm this hypothesis.

### 6.3.2. Physical motion vs. non-physical motion readings

We start the analysis with the distinction “physical” vs. “non-physical motion”. Examples (36–41) present these two readings with each preposition.

(36) *quando eu cheguei na oitava classe* (STP, corpus data, non-physical motion)  
 when I arrived in.the 8<sup>th</sup> grade  
 ‘when I got to the eighth grade’

(37) *quando ela chegou na cidade* (STP, corpus data, physical motion)  
 when she arrived in.the town  
 ‘when she arrived in town’

(38) *uma notícia que chega a alguém* (AP, corpus data, non-physical motion)  
 a news that arrives to someone  
 ‘news that reaches someone’

(39) *e logo que cheguei ao aeroporto* (AP, corpus data, physical motion)  
 and soon that I.arrived to.the airport  
 ‘and as soon as I arrived at the airport’

(40) *vamos agora para gastronomia e culinária angolana*  
 let’s.go now to gastronomy and cuisine Angolan  
 ‘let’s move on now to Angolan gastronomy and cuisine’ (AP, corpus data, non-physical motion)

(41) *quando eu fui para o norte do país* (MP, corpus data, physical motion)  
 when I went to the north of.the country  
 ‘when I went to the north of the country’

Tables 4 and 5 below show the results for each of the two verbs, by preposition, and by AVP.

<i>ir</i>	<i>PPa</i>		<i>PPpara</i>		<i>PPem</i>		total
	non-physical motion	physical motion	non-physical motion	physical motion	non-physical motion	physical motion	
AP	6,83% (14)	22,44% (46)	12,2% (25)	27,8% (57)	1,95% (4)	28,78% (59)	100% (205)
MP	12,53% (57)	42,86% (195)	5,93% (27)	28,79% (131)	1,32% (6)	8,57% (39)	100% (455)
STP	7,58% (30)	25% (99)	19,44% (77)	45,96% (182)	0,25% (1)	1,77% (7)	100% (396)
Total nr.	101	340	129	370	11	105	1056

Table 4: physical vs. non-physical motion with *ir*.

<i>chegar</i>	<i>PPa</i>		<i>PPem</i>		total
	non-physical motion	physical motion	non-physical motion	physical motion	
AP	26,67% (12)	24,44% (11)	11,11% (5)	35,56% (16)	100% (44)
MP	14,29% (17)	13,44% (16)	5,88% (7)	66,39% (79)	100% (119)
STP	27,03% (20)	22,97% (17)	5,41% (4)	44,59% (33)	100% (74)
Total nr.	49	44	16	128	237

Table 5: physical vs. non-physical motion with *chegar*.

The main conclusions can be summarized as follows. In what concerns the occurrences with the verb *ir*, the results are not clear-cut, but:

- (i) both with prepositions *a* and *para*, the AVPs show a regular pattern, as non-physical readings correspond to approximately 1/3 of the physical readings (with *a*: 101 non-physical readings vs. 340 physical readings; with *para*: 129 non-physical readings vs. 370 physical readings); this proportion seems to be related to the physical/non-physical asymmetry, and not to the prepositions.



- (ii) with *em*, there is a sharp contrast, as non-physical readings correspond to approximately 1/10 of the physical readings (11 non-physical readings vs. 105 physical readings).

With the verb *chegar*, the results are slightly different from the ones with *ir*, because, as in EP and BP, preposition *para* does not occur with *chegar*<sup>20</sup>; the comparison between prepositions therefore only regards *a* and *em*:

- (i) with preposition *a*, non-physical readings correspond to approximately the same number of physical readings (a tendency shared among the AVPs): 49 non-physical readings vs. 44 physical readings.
- (ii) with *em* there is again a sharp contrast, as non-physical readings correspond to 1/8 of the physical readings: 16 non-physical readings vs. 128 physical readings.

The detailed results of **Tables 4** and **5** are summarized in **Table 6**, which does not take into account the results by AVP nor by verb.

	<i>a</i>	<i>para</i>	<i>em</i>	total
non-physical motion	49,02% (150)	42,16% (129)	8,82% (27)	100% (306)
physical motion	38,90% (384)	37,49% (370)	23,61% (233)	100% (987)

**Table 6:** physical vs. non-physical motion by preposition in the AVPs under analysis.

As expected, given the variation that characterizes the AVPs, the results are not absolutely clear-cut, but they show a tendency: the use of preposition *em* in predications describing non-physical motion events is substantially reduced compared to the other two prepositions (27 vs. 233 occurrences). On the other hand, prepositions *a* and *para* seem to be chosen irrespective of the distinction “physical/non-physical” motion. Fisher’s test applied to the data in **Table 6** shows that an enormous difference between the pairs *em/a* ( $p < .001$ ), and *em/para* ( $p < .001$ ) can be observed. On the other hand, we found no statistical difference between *para/a* ( $p = 0.440824$ ) regarding the parameter “physical vs. non-physical motion”. Our data therefore seem to confirm Nikitina’s hypothesis: although there is no threshold dividing different uses of prepositions, the predications with directional *em* are typically those that correspond to events describing some change in the location of an entity that can be identified by spatial coordinates. On the contrary, if predications express non-physical motion (that is, events of change along a scale, or states denoting a location), prepositions *a* and *para* are typically used.

---

<sup>20</sup> Only one occurrence in the data.

### 6.3.3. Semantics of the preposition's complement

In this section, we present the results concerning the semantic type of NP, that is, if it is a container or an area. The data concerning the cases marked as “undefined” (see 6.3.1 above) were removed, because they don't add any relevant information to our analysis. Moreover, we only analyzed predications marked as “physical” motion (cf. previous section). Corpus-based examples (42–45), with *em*, illustrate the different combinations.

- (42) *fui em casa de meu pai* (MP, corpus data, *ir* + container)  
I.went in house of my father  
'I went to my father's house'
- (43) *quimbanda é aquela pessoa que vai na mata* (MP, corpus data, *ir* + area)  
*quimbanda* is that person that goes in.the woods  
'Quimbanda is that person who goes in the woods'
- (44) *quando chego no hospital* (STP, corpus data, *chegar* + container)  
when I.arrive in.the hospital  
'when I arrive at the hospital'
- (45) *estávamos a chegar na zona* (MP, corpus data, *chegar* + area)  
we.were at arrive in.the area  
'we were arriving in the area'

Tables 7, 8, and 9 present the results according to preposition (*em*, *a*, *para*), except for the combination preposition *para* + verb *chegar* for which we only found one occurrence (cf. fn. 20 above), which we did not include in Table 9.

<i>ir &amp; chegar</i> PPem	Container	Area	total
AP	52,54% (31)	47,46% (28)	100% (59)
MP	72,38% (76)	27,62% (29)	100% (105)
STP	77,50% (31)	22,50% (9)	100% (40)
Total nr.	138	66	204

Table 7: Verbs *ir & chegar* + preposition *em*.

<i>ir &amp; chegar</i> PPa	Container	Area	total
AP	53,33% (24)	46,67% (21)	100% (45)
MP	62,43% (113)	37,57% (68)	100% (181)
STP	36,54% (38)	63,46% (66)	100% (104)
Total nr.	175	155	330

**Table 8:** Verbs *ir & chegar* + preposition *a*.

<i>Ir</i> PPpara	Container	Area	total
AP	33,96% (18)	66,04% (35)	100% (53)
MP	48,76% (59)	51,24% (62)	100% (121)
STP	50,31% (81)	49,69% (80)	100% (161)
Total nr.	158	177	335

**Table 9:** Verb *ir* + preposition *para*.

The results are not clear-cut, but they show that preposition *em* (Table 7) is more prone to occur with containers than prepositions *a* and *para* (Tables 8 and 9), particularly in MP (72,38%) and STP (77,50%) but less so in AP (52,54%).

With preposition *a*, in Table 8, the overall difference between containers and areas is small (175 vs. 155 occurrences): MP confirms its preference for containers with this preposition (62,43%); STP, on the other hand, shows a preference for areas (63,46%). Finally, with preposition *para* (and verb *ir* only), in Table 9, the overall difference between containers and areas is, once again, small, but in this case the preference leans slightly toward areas (158 vs. 177 occurrences), a tendency which is clearer in AP (66,04%).

The detailed results of Tables 7, 8 and 9 are summarized in Table 10, which does not take into account the results by AVP.

	<i>em</i>	<i>a</i>	<i>para</i>	total
container	29,30% (138)	37,15% (175)	33,55% (158)	100% (471)
area	16,58% (66)	38,95% (155)	44,47% (177)	100% (398)

**Table 10:** Containers and areas by preposition in the AVPs under analysis.

This table shows the contrast between the preference for containers instead of areas with Goal phrases headed by *em* in AVPs (29,30% vs. 16,58%). With respect to PPs headed by the prepositions *a* and *para*, the difference between these two types of NP is not clear-cut, with a slight preference of *para* toward areas (44,47%). In fact, regarding the parameter “container vs. area” in **Table 10**, Fisher’s test shows the same tendencies noticed before (concerning the parameter “physical vs. non-physical motion”, in **Table 6**), i.e., statistical differences can be observed between the pairs *em/a* ( $p = 0.001$ ) and *em/para* ( $p < .001$ ), but not between *para/a* ( $p = 0.14082$ ). In the next section it will be shown that the tendencies pinpointed in this section are not identical for the verbs under analysis.

#### 6.3.4. Semantics of the verb

Recall that, in Nikitina’s (2008) proposal, the lexical information exhibited by the verb form is a relevant factor: locative PPs with Goal interpretation are found more often with non-durative verbs than with durative ones. We mentioned previously that Portuguese *chegar* behaves as a non-durative verb, whereas *ir* is ambiguous between durative and non-durative readings. This difference between these two verbs seems to explain the different results described earlier: since *chegar* is a non-durative verb, it combines more often with directional *em* phrases in all three varieties at stake. But the fact that *ir* is ambiguous between durative and non-durative readings does not favor the directional readings of Goals headed by *em*. These conclusions seem to be reinforced when we look at the data that consider the distribution of examples using the parameter “type of NP”, shown in **Tables 11** and **12** below.

The detailed results of **Tables 11** and **12** are summarized in **Table 13**, which does not take into account the results by AVP.

These results clearly show that, with *chegar*, a non-durative verb, Goal phrases are typically introduced by *em* (75,16%). A sharp contrast emerges with *ir*, a verb that lexically allows both durative and non-durative readings, where the occurrences of Goal phrases introduced by *em* are

substantially reduced (11,72%).<sup>21</sup> This does not mean that the type of NP is not relevant: on the contrary, it was shown that with both verbs there are more containers than areas occurring in Goal phrases introduced by *em*. However, these results show that the variable type of NP appears to be less influential than the variable type of verb. Since only two verbs were analyzed, these conclusions must be confirmed by an analysis that encompasses other verbs of movement.

<i>ir</i>	<b>PPem</b>		<b>PPa</b>		<b>PPpara</b>		<b>total</b>
	<b>Container</b>	<b>Area</b>	<b>Container</b>	<b>Area</b>	<b>Container</b>	<b>Area</b>	
AP	15,38% (20)	17,69% (23)	13,85% (18)	12,31% (16)	13,85% (18)	26,92% (35)	100% (130)
MP	7,50% (24)	2,81% (9)	31,88% (102)	20,00% (64)	18,44% (59)	19,37% (62)	100% (320)
STP	1,16% (3)	1,55% (4)	11,63% (30)	23,26% (60)	31,39% (81)	31,01% (80)	100% (258)
Total nr.	47	36	150	140	158	177	708

**Table 11:** NP containers and areas with the verb *ir*.

<i>chegar</i>	<b>PPem</b>		<b>PPa</b>		<b>total</b>
	<b>Container</b>	<b>Area</b>	<b>Container</b>	<b>Area</b>	
AP	40,74% (11)	18,52% (5)	22,22% (6)	18,52% (5)	100% (27)
MP	59,77% (52)	22,99% (20)	12,64% (11)	4,60% (4)	100% (87)
STP	59,57% (28)	10,64% (5)	17,02% (8)	12,77% (6)	100% (47)
Total nr.	91	30	25	15	161

**Table 12:** NP containers and areas with the verb *chegar*.

<sup>21</sup> Applying Fisher's test to the data in Table 13 shows that remarkable differences between all pairs of prepositions can be observed. But notice that the existence of a statistical difference between *para* and *a* is biased by the fact that only the latter preposition occurs with the verb *chegar*.

	<i>em</i>	<i>a</i>	<i>para</i>	<i>total</i>
<i>ir</i>	11,72% (83)	40,96% (290)	47,32% (335)	100% (708)
<i>chegar</i>	75,16% 121	24,84% 40	0% 0	100% (161)

**Table 13:** Verbs *ir* and *chegar* by preposition in the AVPs under analysis.

Despite these limitations, these preliminary results lend overall support to Nikitina’s proposal that one cannot point out one single factor underlying the occurrence of Goal phrases headed by *em*. Instead, it seems that there are several factors with a different weight that contribute to the observed tendencies: (i) the semantic nature of the NP complement is a relevant factor (containers favor the occurrence of directional *em* phrases), but (ii) the semantics of the verb is more relevant than the semantics of the NP (*chegar*, a non-durative verb, favors the occurrence of directional *em* phrases).

In fact, the application of the “Joint Entropy” (JE) and “Mutual Information” (MI) functions to the data consistently reveals this correlation, as shown in **Table 14**. These functions (Manning & Schutze, 1999) allow us to quantify the degree of association between two variables. For example, if we have two variables potentially contributing to the ranking/classification of a third one, we can measure which one contributes the most (being the most relevant). The JE measures a quantity having an opposite sense of the MI. JE measures the level of irrelevance of one variable with respect to another one, and the lower the value the more dependence/order exists. On the contrary, MI measures how much the phenomenon that jointly involves the two variables is beyond random, meaning that there is information in the observed association that can be measured in Shannon bits.<sup>22</sup> In this case, the higher the MI, the greater the degree of association between the variables. The application of both functions to the data (including the three prepositions at stake) converges toward the same result, as shown in **Table 14**: the most relevant parameter is the verb (the highest MI and the lowest JE) and the least relevant parameter is the opposition “container vs. area” (the lowest MI and the highest JE), with the parameter “physical vs. non-physical motion” being in between.

---

<sup>22</sup> The Shannon bit is a unit of information related to the conventional bit (used in computers). While in the latter the value is always an integer (e.g., 2 bits, 64 bits), in the former the value tends not to be (e.g., 0.26 bits, 3.124 bits), resulting from the use of the base 2 logarithm for its calculation. A certain amount of Shannon bits represents the combined number of bits necessary to codify a number of random events with a certain probability with respect to each event.

Parameters		Functions	
		Joint Entropy	Mutual Information
Verb	<i>ir</i>	1,98261	0,2602872
	<i>chegar</i>		
Motion	non-physical	2,29107	0,0207105
	physical		
NP Goal complement	container	2,52796	0,0183608
	area		

**Table 14:** Parameters relevance using “Joint Entropy” and “Mutual Information”.

#### 6.4. Summary

To conclude, the data we presented confirm Nikitina’s (2008) hypothesis. Although there is no threshold dividing different uses of prepositions, the following general picture emerges. The use of Goal phrases headed by *em* in AVPs is favored by semantic factors with different weights:

- (i) The lexical semantics of the verb: non-durative verb *chegar* favors *em* Goal phrases (contrary to *ir*, which licenses both durative and non-durative readings).
- (ii) The type of eventuality described: “physical” motion events favor *em* Goal phrases (instead of metaphorical motion events or fictive motion situations).
- (iii) The type of NP complement of the preposition: containers favor *em* Goal phrases (instead of areas).

According to our data, (i) appears to play a primary role, (iii) is the least relevant factor, whereas (ii) is somewhere in between. However, further studies that include other verbs of movement are required to confirm this claim.

All in all, we conclude that, typically, with *em* Goal phrases, the predications correspond to non-durative physical events of movement, i.e., events describing some change in the location of an entity (a change that can be identified by spatial coordinates), and, in the end of the events, the entity is located typically inside the region associated to the entity denoted by the NP complement of the preposition. A semantic account for the occurrence of Goal *em* phrases could therefore explain the AVP data in the same way Nikitina (2008) and Kopecka (2009), for instance, account for the English and French data. Furthermore, this analysis is aligned with the analyses of BP data previously referred. In fact, there is nothing peculiar about the fact that

AVPs allow Goal phrases introduced by *em*. Being a crosslinguistically widespread pattern, AVPs behave just like many other languages. EP, on the contrary, constitutes an exception to this regularity since this pattern is restricted to a few verbs only.<sup>23</sup> In sum, we argue that general semantic (and pragmatic) factors must be taken into consideration in the analysis of AVPs, namely when scrutinizing the importance of language contact to explain the occurrence of Goal phrases introduced by *em*.

## 7. Conclusions

The main goal of this paper was to assess the use of preposition *em* introducing Goal phrases in urban AVPs, in order to test the hypothesis of language contact as the major cause for this use. To do so, we analyzed examples extracted from spoken corpora of three AVPs (AP, MP, and STP) with two inherently directed motion verbs *ir* ‘to go’ and *chegar* ‘to arrive’ and we compared the strategies to express Goal phrases in these AVPs with the ones used in the main contact languages (Kimbundu for AP; Changana for MP; and Forro for STP).

In general, we found that AVPs typically use PPs as Goal phrases and that there is a difference between the verbs: *ir* generally occurs with the standard EP strategies (*a* and *para*), whereas *chegar* exhibits strong variation between the standard EP strategy with *a* and the non-standard EP strategy with *em*. In addition, STP stands out in comparison to AP and MP because of a strong tendency for *ir* to select *para* and due to the fact that this variety also exhibits a fair number of cases where *chegar* selects an NP argument.

It was further shown that the language contact hypothesis lacks explanatory power with respect to the use of prepositions with the verbs *ir* and *chegar* across AVPs. The three main contact languages exhibit internally uniform patterns for the verbs ‘to go’ and ‘to arrive’ (PPs headed by *ku* in Kimbundu; morphological strategies in Changana; and NPs in Forro), which do not match the patterns in the respective AVPs, where PPs are the dominant pattern and *ir* and *chegar* exhibit different patterns of variation. Although the semantic hypothesis was not explicitly applied to the contact languages discussed in section 4, they lack the kind of sensitivity to the semantics of the predicate and to the nature of the Goal (area or container) observed in the AVPs with regard to the verbs of directed motion at stake. At best, there is only a very mild correlation between some of the patterns that occur in the AVPs and in the contact languages. In addition, we argued in section 5 that the contact-induced morphological incorporation hypothesis of preposition *em* inspired on the Bantu locative classes, originally proposed for MP, lacks empirical motivation.

Given the weaknesses of the contact-induced hypothesis, we put forward a different hypothesis. We assumed, as in Nikitina (2008), that the use of locative prepositions heading

---

<sup>23</sup> Although the scope of this paper is not EP, this language belongs typologically to the group of languages in which Location and Goal receive differentiated (prepositional) marking, whereas (spoken) AVPs, as well as BP, can receive the same marking for these two semantic roles.



Goal phrases can be explained taking into consideration semantic/pragmatic factors. To test this hypothesis, we analyzed the data using the following parameters: the semantic nature of the predication, i.e., whether it describes an event of movement that can be traced using spatial coordinates; the semantic nature of the complement of the preposition, i.e., whether the NP denotes an entity with well-defined boundaries; and the lexical semantics of both verbs, i.e., whether the verb describes a durative or non-durative change of place.

The main findings of this approach are as follows. In the first place, and as expected, there are no clear-cut rules regulating the use of *em* in AVPs. Instead, directional readings of preposition *em* in AVPs are favored by the following properties, where i) was shown to be the most prominent and (iii) the least relevant:

- (i) the predication corresponds to a non-durative change of place.
- (ii) the speaker wants to describe an event in which there is an entity that moves toward a reference object, and this movement can be described by a series of spatial coordinates (i.e, a movement in the “real world”).
- (iii) the reference object describing the final location of the mover has well-defined boundaries (containers).

The main conclusion that emerges is that in the case of the contemporary, urban AVPs the contact-induced hypothesis is seriously flawed by the empirical data, whereas the occurrence of Goal phrases headed by locative prepositions (such as *em*) is attested in different, typologically unrelated languages. In other words, we are dealing with a cross-linguistic phenomenon that relies on the mechanisms available in each language to express motion and on general semantic and pragmatic principles.

There are, however, some shortcomings of our analysis that require future research, in particular the amount of analyzed verbs – only two. Different types of verbs of movement (durative and non-durative verbs; manner of motion and inherently directed motion verbs) will have to be added to the sample in order to test if our general conclusions hold regarding the durative vs. non-durative nature of (the lexical semantics of) the verb and its relative weight in the licensing of Goal phrases headed by *em*. Other aspects of Nikitina’s proposal, namely the hypothesis that Goal-*em* phrases cannot cooccur in the same predication together with Source phrases or with verbs that describe highly specific manners of motion (e.g., to crawl), will be difficult to evaluate, because they occur only scarcely in the spoken corpora. A workaround would be to carry experimental tasks with native speakers of the AVPs. The increase in the number of contexts will also allow us to further assess whether the parameter “type of NP” is indeed the less influential factor in determining the occurrence of Goal phrases headed by *em* in AVPs.

---

## Abbreviations

AP = Angolan Portuguese; APPL = applicative; AVP = African variety of Portuguese; BP = Brazilian Portuguese; CONN = connective; DEM = demonstrative; EP = European Portuguese; FRM = formal; FUT = future; LOC = locative; MP = Mozambican Portuguese; NP = noun phrase; POSS = possessive; PP = preposition phrase; PR = present; PT = past tense; STP = Santomean Portuguese; V = verb; VF = final vowel.

## Acknowledgements

The research in this paper was developed within project PALMA (Possession and location: microvariation in African varieties of Portuguese, PTDC/LLT-LIN/29552/), funded by the Portuguese Foundation for Science and Technology (FCT). We are greatly indebted to Afonso Miguel and Feliciano Chibutane for their assistance with the Kimbundu and Changana data, to Rita Gonçalves, who participated in an early version of this research, to three anonymous reviewers whose comments were crucial in improving an earlier version of this paper, and to the editorial team of JPL for their tireless support throughout the whole process.

## Competing Interests

The authors have no competing interests to declare.

---

## References

- Adriano, P. (2014). *Tratamento morfossintático de expressões e estruturas frásicas do Português em Angola. Divergências em relação à norma europeia* [Morphosyntactic treatment of sentence structure and expressions of Portuguese in Angola. Divergences with respect to the European standard] (Unpublished doctoral dissertation). University of Évora, Portugal. [https://dspace.uevora.pt/rdpc/bitstream/10174/11802/1/Doutoramento\\_Paulino%20Adriano.pdf](https://dspace.uevora.pt/rdpc/bitstream/10174/11802/1/Doutoramento_Paulino%20Adriano.pdf)
- Alonge, A. (1997). Semantica lessicale e proprietà sintattiche dei verbi di movimento italiani: analisi di dati acquisiti da dizionari di macchina e da un corpus testuale computerizzato [Lexical semantics and syntactic properties of Italian verbs of movement: analysis of data from machine dictionaries and from a computerized textual corpus]. *Atti del III Convegno della Società Internazionale di Linguistica e Filologia Italiana*, 31–63.
- Araújo, P. (2013). *Domínios conceituais das construções locativas, existenciais, comitativas e possessivas em línguas bantas* [Conceptual domains of locatives, existentials, comitatives, and possessives in Bantu languages](Unpublished doctoral Dissertation). University of São Paulo, Brazil. <https://www.teses.usp.br/teses/disponiveis/8/8139/tde-28082013-153034/pt-br.php>
- Avelar, J. (2017). Complementos direcionais em afro-variedades de português e espanhol [Directional complements in afro-varieties of Portuguese and Spanish]. *Moderna språk*, 111(2), 15–44. <https://ojs.ub.gu.se/index.php/modernasprak/article/view/4173>

- Avelar, J., & Álvarez López, L. (2018). Directional complements, existential sentences and locatives in the Afro-Brazilian continuum of Portuguese. In L. Álvarez, López, P. Gonçalves & J. Avelar (Eds.), *The Portuguese language continuum in Africa and Brazil* (pp. 185–210). Amsterdam: John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/ihll.20.09dea>
- Beavers, J., Levin, B., & Tham, S. (2010). The typology of motion expressions revisited. *Journal of Linguistics*, 46(2), 331–377. DOI: <https://doi.org/10.1017/S0022226709990272>
- Blomberg, J., & Zlatev, J. (2014). Actual and non-actual motion: why experimentalist semantics needs phenomenology (and vice versa). *Phenomenology and the cognitive sciences*, 13(3), 395–418. DOI: <https://doi.org/10.1007/s11097-013-9299-x>
- Cabral, L. (2005). *Complementos verbais preposicionados do português em Angola* [Prepositioned verbal complements of Portuguese in Angola] (Unpublished master's thesis). University of Lisbon, Portugal. <http://hdl.handle.net/10451/4278>
- Chatelain, H. (1888–1889). *Grammatica elementar do kimbundu ou lingua de Angola* [Elementary grammar of Kimbundu or language of Angola]. Ridgewood, New Jersey: The Gregg Press Incorporated.
- Chavagne, J.-P. (2005). *La langue portugaise d'Angola: Étude des écarts par rapport à la norme européenne du portugais* [The Portuguese language in Angola: study of the differences with respect to the European standard] (Unpublished doctoral dissertation). University Lumière, France.
- Chimbutane, F. (2002). *Grammatical functions in Changana: types, properties and function alternations* (Unpublished master's thesis). The Australian National University, Australia. DOI: <https://doi.org/10.25911/5d515645659ae>
- Diarra, B. (1990). *Gramática kimbundu* [Kimbundu grammar]. Luanda: Secretary of State for Culture, Institute of National Languages, Angola.
- Dowty, D. (1979). *Word meaning and Montague grammar. The semantic of verbs and times in generative semantics and in Montague's PTQ*. Dordrecht: Springer Dordrecht. DOI: <https://doi.org/10.1007/978-94-009-9473-7>
- Eddington, D. (2016). *Statistics for linguists: A step-by-step guide for novices*. Cambridge: Cambridge Scholars Publishing.
- Fábregas, A. (2007). An exhaustive lexicalisation account of directional complements. *Tromsø Working Papers on Language & Linguistics*, 34(2), 165–199. DOI: <https://doi.org/10.7557/12.110>
- Farias, J. G. de (2006). Variação entre *a*, *para* e *em* no português brasileiro e no português europeu: algumas notas [Variação between *a*, *para* and *em* in Brazilian Portuguese and European Portuguese: some notes]. *Letras de Hoje*, 41(1), 213–234.
- Ferreira, T. L., & Basso, R. M. (2019). Preposições de Alvo no português brasileiro: uma comparação entre 'para' e 'até' [Goal prepositions in Brazilian Portuguese: a comparison between *para* and *até*]. *Revista Linguística*, 15(3), 43–68. DOI: <https://doi.org/10.31513/linguistica.2019.v15n3a27505>
- Firmino, G. (2002). A “questão linguística” na África pós-colonial: o caso do português e das línguas autóctones em Moçambique [The “language issue” in postcolonial Africa: the case of Portuguese

- and the autochthonous languages in Mozambique]. Maputo: Promédia. DOI: <https://doi.org/10.1590/1981-5794-e12448>
- Folli, R., & Ramchand, G. (2005). Prepositions and results in Italian and English: an analysis from event decomposition. In H. Verkuyl, H. de Swart & A. van Hout (Eds.), *Perspectives on aspect* (pp. 81–105). Dordrecht: Kluwer. DOI: [https://doi.org/10.1007/1-4020-3232-3\\_5](https://doi.org/10.1007/1-4020-3232-3_5)
- Gehrke, B. (2007). Putting path in place. In E. Puig-Waldmüller (Ed.), *Sinn und Bedeutung 11* (pp. 244–260). Barcelona: Universitat Pompeu Fabra. DOI: <https://doi.org/10.18148/sub/2007.v11i0.643>
- Gehrke, B. (2008). *Ps in motion: on the semantics and syntax of P elements and motion events (LOT Dissertation Series)*. Utrecht: LOT. [https://www.lotpublications.nl/Documents/184\\_fulltext.pdf](https://www.lotpublications.nl/Documents/184_fulltext.pdf)
- Gonçalves, P. (2002). The role of ambiguity in second language change: the case of Mozambican African Portuguese. *Second Language Research*, 18 (4), 325–347. DOI: <https://doi.org/10.1191/0267658302sr209oa>
- Gonçalves, P. (2004). Towards a unified vision of classes of language acquisition and change: arguments from the genesis of Mozambican African Portuguese. *Journal of Pidgin and Creole Languages*, 19(2), 225–259. DOI: <https://doi.org/10.1075/jpcl.19.2.01gon>
- Gonçalves, P. (2005). A formação de variedades africanas do português: argumentos para uma abordagem multidimensional [The formation of African varieties of Portuguese: arguments in favor of a multidimensional approach]. *A língua portuguesa: presente e futuro* (pp. 223–242). Lisbon: Calouste Gulbenkian Foundation.
- Gonçalves, P. (2010). *A génese do Português de Moçambique* [The genesis of Mozambican Portuguese]. Lisbon: INCM.
- Gonçalves, P. (2013). O português em África [Portuguese in Africa]. In E. P. Raposo, M. F. B. do Nascimento, M. A. C. da Mota, L. Segura & A. Mendes (Orgs.), *Gramática do português* (Vol. 1, pp. 157–178). Lisbon: Calouste Gulbenkian Foundation.
- Gonçalves, P., & Chimbutane, F. (2004). O papel das línguas bantu na génese do português de Moçambique: o comportamento sintáctico de constituintes locativos e direccionais [The role of the Bantu languages in the genesis of Mozambican Portuguese: the syntactic behavior of locative and directional complements]. *Papia*, 14, 7–30.
- Gonçalves, R. (2010). *Propriedades de subcategorização verbal no português de São Tomé* [Verbal subcategorization properties in Santomean Portuguese] (Unpublished master's thesis). University of Lisbon, Lisbon. <http://hdl.handle.net/10451/4252>
- Gonçalves, R., Duarte, I., & Hagemeyer, T. (2022). Dative microvariation in African varieties of Portuguese. *Journal of Portuguese Linguistics*, 21:6, pp. 1–39. DOI: <https://doi.org/10.16995/jpl.8488>
- Gonçalves, R., Hagemeyer, T., Alcântara, C., Cornejo, C., Madureira, R., Génereux, M., & Mendes, A. (2021). *PALMA Corpus São Tomé e Príncipe*. Lisbon: Center of Linguistics of the University of Lisbon.
- Hagemeyer, T. (2004). Going in the clause: ba and be in Santome. *Journal of Portuguese Linguistics*, 3(2), 71–95. DOI: <https://doi.org/10.5334/jpl.12>

- Hagemeijer, T. (2016). O português em contacto em África [Portuguese in contact in Africa]. In A. M. Martins & E. Carrilho (Eds.), *Manual de Linguística Portuguesa* (pp. 43–67). Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110368840-004>
- Hagemeijer, T., Amaro, H., Génereux, M, Hendrickx, I., Mendes, A., & Tiny, A. (2014). *Santome corpus*. Lisbon: Center of Linguistics of the University of Lisbon.
- Hagemeijer, T., Chimbutane, F., Gonçalves, R., Leal, A., & Miguel, A. (2019). A expressão de alvo e origem de verbos de movimento em variedades africanas do português [The expression of goal and source of verbs of movement in African varieties of Portuguese]. Paper presented at the XXXV Encontro da Associação Portuguesa de Linguística, Universidade do Minho.
- Hagemeijer, T., Madureira, R., Cornejo, C., Justino, V., Campos, M., Gonçalves, R., Génereux, M., & Mendes, A. (2021). *PALMA Corpus Moçambique*. Lisbon: Center of Linguistics of the University of Lisbon.
- Hagemeijer, T., Mendes, A., Gonçalves, R., Cornejo, C., Madureira, R. & Génereux, M. (2022). The PALMA corpora of African varieties of Portuguese. In N. Calzolari, F. Béchet, Ph. Blache, Kh. Choukri, Ch. Cieri, Th. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odiijk, S. Piperidis (Eds.), *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)* (pp. 5047–5053). Paris: European Language Resources Association. [www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.539.pdf](http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.539.pdf)
- INE (Instituto Nacional de Estatística de Angola) (2016). *Resultados definitivos do recenseamento geral da população e da habitação de Angola 2014* [Final results of the general population and housing census of Angola 2014].
- INE (Instituto Nacional de Estatística de Moçambique) (2019). *IV Recenseamento geral da população e habitação 2017* [IV general population and housing census 2017].
- INE (Instituto Nacional de Estatística de São Tomé e Príncipe) (2013). *IV Recenseamento geral da população e da habitação 2012. Características educacionais da população* [IV general population and housing census 2012. Educational characteristics of the population].
- Kopecka, A. (2009). L'expression du déplacement en français: l'interaction des facteurs sémantiques, aspectuels et pragmatiques dans la construction du sens spatial [The expression of movement in French: the interaction of semantic, aspectual, and pragmatic factors in spatial constructions]. *Langages*, 173, 54–75. DOI: <https://doi.org/10.3917/lang.173.0054>
- Langacker R. W. (1986). Abstract motion. In V. Nikiforidou, M. Van Clay, M. Niepokuj & D. Feder, *Proceedings of the 12th Annual Meeting of the Berkeley Linguistics Society* (pp. 455–471). Berkeley, CA: Berkeley Linguistics Society. DOI: <https://doi.org/10.3765/bls.v12i0.3317>
- Leal, A. (2009). *Semântica aspectual e nominal. Contributo das expressões nominais para a construção aspectual das frases* [Aspectual and nominal semantics: Contribution of the nominal expressions to the aspectual construction of sentences] (Unpublished doctoral dissertation). University of Porto, Porto.
- Leal, A., Oliveira, F., & Silvano, P. (2017). Verbos de movimento e preposições direcionais [Verbs of movement and directional prepositions]. *Revista da Associação Portuguesa de Linguística*, 3, 119–133. DOI: <https://doi.org/10.26334/2183-9077/rapln3ano2017a8>

- Leal, A., Oliveira, F., & Silvano, P. (2018). Path Scales: directed-motion verbs, prepositions and telicity in European Portuguese. In D. Ayoun, A. Celle, & L. Lansari (Eds.), *Tense, aspect, modality and evidentiality: cross-linguistic perspectives* (pp. 335–354). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/slcs.197>
- Levin, B. (1993). *English verb classes and alternations: a preliminary investigation*. Chicago, IL: The University of Chicago Press.
- Levin, B., Beavers, J., & Tham, S. (2009). Manner of motion roots across languages: Same or different? Paper presented at *Roots: Word Formation from the Perspective of 'Core Lexical Elements'*, University of Stuttgart.
- Lucchesi, D. (2001). As duas grandes vertentes da história sociolinguística do Brasil (1500–2000) [The two main aspects of Brazil's sociolinguistic history (1500–2000)]. *DELTA*, 17(1), 97–130. DOI: <https://doi.org/10.1590/S0102-44502001000100005>
- Manning, C., & Schutze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.
- Marques, I. (1985). *Algumas considerações sobre problemática linguística em Angola* [Some considerations on Angola's linguistic problematic]. Luanda: INALD.
- Marten, L. (2010). The great siSwati locative shift. In A. Breitbarth, C. Lucas, S. Watts & D. Willis (Eds.), *Continuity and change in grammar* (pp. 249–267). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.159.12mar>
- Matsumoto, Y. (1996). Subjective motion and English and Japanese verbs. *Cognitive Linguistics*, 7(2), 183–226. DOI: <https://doi.org/10.1515/cogl.1996.7.2.183>
- Miguel, A. (2019). *Integração morfológica e fonológica de empréstimos lexicais bantos no português oral de Luanda* [Morphological and phonological integration of Bantu loanwords in spoken Luanda Portuguese] (Unpublished doctoral dissertation). School of Arts and Humanities of University of Lisbon, Lisbon. <http://hdl.handle.net/10451/37919>
- Miguel, A., Cornejo, C., Madureira, R., Silva, D., Hagemeyer, T., Gonçalves, R., Génereux, M., & Mendes, A. (2021). *PALMA Corpus Angola*. Lisbon: Center of Linguistics of the University of Lisbon.
- Mingas, A. (2000). *Interferência do kimbundu no português falado em Lwanda* [Interference of Kimbundu on Portuguese spoken in Luanda]. Porto: Campo das Letras.
- Moens, M. (1987). *Tense, aspect and temporal reference* (Unpublished doctoral dissertation). Edinburgh: University of Edinburgh. <http://hdl.handle.net/1842/5369>
- Mollica, M. (1996). A regência verbal do verbo *ir* em movimento [The verbal government of verb *ir* in movement]. *Padrões sociolinguísticos*, 85–119. Rio de Janeiro: Tempo Brasileiro.
- Ngunga, A., & Simbine, M. (2012). *Gramática descritiva da língua changana* [Descriptive grammar of Changana]. Maputo: Center of African Studies, University Eduardo Mondlane.
- Nikitina, T. (2008). Pragmatic factors and variation in the expression of spatial goals. The case of into vs. in. In A. Asbury, J. Dotlačil, B. Gehrke & R. Nouwen (Eds.), *Syntax and semantics of spatial P* (pp. 175–195). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.120.09nik>

- Nikitina, T. (2017). Ablative and allative marking of static locations. A historical perspective. In S. Luraghi, T. Nikitina & Ch. Zanchi (Eds.), *Space in diachrony. Studies in Language Companion Series* (pp. 67–94). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/slcs.188.03nik>
- Nikitina, T., & Maslov, B. (2013). Redefining *constructio praegnans*: on the variation between allative and locative expressions in Ancient Greek. *Journal of Greek Linguistics*, 13(1), 105–142. DOI: <https://doi.org/10.1163/15699846-13130107>
- Oliveira, F., Cunha, L.F., Leal, A., Silvano, P., Silva, F., & Ferreira, I. (2021). Para uma análise semântica de *para* locativo em construções estativas [Toward a semantic analysis of locative *para* in stative constructions]. *Linguística. Revista de Estudos Linguísticos da Universidade do Porto* (Special issue in Honorem Ana Maria Barros de Brito), 259–279. DOI: <https://doi.org/10.21747/16466195/lingespa12>
- Pantcheva, M. (2010). The syntactic structure of locations, goals, and sources. *Linguistics*, 48(5), 1043–1081. DOI: <https://doi.org/10.1515/ling.2010.034>
- Rammé, V. (2017). *Mudança semântica no PB: reanálise restringida pela hierarquia funcional-conceitual universal* [Semantic change in Brazilian Portuguese: reanalysis restricted by the universal functional-conceptual hierarchy] (Unpublished doctoral dissertation). Federal University of Paraná, Brazil. <https://hdl.handle.net/1884/47405>
- Raposo, E.P., & Xavier, M.F. (2013). Preposição e sintagma preposicional [Preposition and prepositional phrase]. In E. Raposo, M. F. Nascimento, M. A. Mota, L. Segura & A. Mendes (Orgs.), *Gramática do português* (Vol. II, pp. 1497–1568). Lisbon: Calouste Gulbenkian Foundation.
- Sitoe, B. (2001). *Verbs of motion in Changana*. Leiden: CNWS. <https://lib.ugent.be/catalog/rug01:000699182>
- Stroud, C. (1997). O corpus: antecedentes, quadro teórico e aspirações práticas [The corpus: antecedents, theoretical framework, and practical aspirations]. In C. Stroud & P. Gonçalves (Orgs.), *Panorama do português oral de Maputo. Volume I – Objectivos e métodos* (pp. 11–45). Maputo: INDE.
- Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, MA.: MIT Press. DOI: <https://doi.org/10.7551/mitpress/6847.001.0001>
- Thomas, E. (2004). On ‘syntactic’ versus ‘semantic’ telicity: evidence from *in* and *on*. *Belgian Journal of Linguistics*, 18, 145–166. DOI: <https://doi.org/10.1075/bjl.18.08tho>
- Ursini, F.-A. (2013). On the syntax and semantics of Spanish spatial prepositions. *Borealis: An International Journal of Hispanic Linguistics*, 2(1), 117–166. DOI: <https://doi.org/10.7557/1.2.1.2403>
- Vendler, Z. (1957). Verbs and times. *Philosophical Review*, 56, 143–160. DOI: <https://doi.org/10.2307/2182371>
- Vieira, M. (2009). Variação das preposições em verbos de movimento [Variation of prepositions with movement verbs]. *Signum: Estudos da Linguagem*, 12(1), 423–445. DOI: <https://doi.org/10.5433/2237-4876.2009v12n1p423>

Wiedemer, M. (2008). *A regência variável do verbo ir de movimento na fala de Santa Catarina* [The variable government of the verb of movement *ir* in Santa Catarina's speech] (Unpublished master's thesis). Federal University of Santa Catarina, Brazil. <http://repositorio.ufsc.br/xmlui/handle/123456789/91439>

Zeller, J. (in press). Locatives. In L. Marten, N. Kula, E. Hurst & J. Zeller (Eds.), *The Oxford guide to the Bantu languages*. Oxford: Oxford University Press. [www.jzeller.de/pdf/Loc.pdf](http://www.jzeller.de/pdf/Loc.pdf)

