Modelling time and space in Brazilian culture

Daniel Couto-Vale
RWTH Aachen
daniel.couto-vale@ifaar.rwth-aachen.de

Rodrigo de Oliveira
University of Aberdeen
rodrigodeoliveira@abdn.ac.uk

Abstract

In order to analyze and synthesize spatial language in Brazilian Portuguese automatically, a machine needs a linguistic model that fits the sort of wordings that adult Brazilians express and can understand. In this paper, we analyzed a corpus of spatial actions and relations manually using the categories of the Generalized Upper Model (GUM) and verified how far this linguistic model covered the instances of our corpus. We found uncovered spatial relations and contextualization constraints and, based on these findings, we proposed (a) a reformulation of GUM's typology for relational figures as well as (b) the notion of relational stability as a culture-specific contextualization constraint.

1 Introduction

Imagine you are in your flat sitting on an autonomous wheelchair, which you can control via voice commands. After someone knocks on the entrance door, you say to the wheelchair: "Take me to the door." The expected reply by the wheelchair would be: "OK. I'll take you there." Full understanding can only be established in such a way if the wheelchair is able to use context to infer which door you meant. Since the flat has many doors (e.g. to the toilet, living room, etc.) the nominal group "the door" represents the most relevant door in the context of situation. Likewise, the reply by the wheelchair omits the type of thing where it is taking you, and it is your job to understand that "there" refers to a position relative to the door. If we wish to allow humans to speak to machines like we do to other humans, we need to model what kinds of phenomena are represented by utterances first, before we can recognize the relevant ones in context.

Modeling the content of an utterance is the domain of semantics while relating this content to context of the utterance is the domain of pragmatics. Such a linguistic model can be reused across different contexts as it is meant to be used in the interface between semantics and pragmatics, i.e. before the contextualization step during understanding and after the sentence planning step during verbalization. Therefore it must make the minimum meaning commitment to avoid ambiguity in linguistic analysis, while being specific enough to specify a unique grammatical unit for linguistic synthesis.

One specific use of language is to describe the location of things; this manifestation of language has been labelled *spatial language*. In modelling spatial language, geometrical accounts have been the predominant first choice of researchers (Herskovits, 1980; Talmy, 1983; Kracht, 2002). The second choice has been to use relaxed geometrical rules (Talmy, 1983; Herskovits, 1985; Kracht, 2002; Francez and Steedman, 2006). However, Zwarts (2005) and Bateman (2010) show that language commits to qualitative and functional notions of space that are independent of time and three-dimensional regions. With a functional approach, the semantic contribution of spatial terms is formalized as an intermediary constraint for identifying entities and relations in a situation and not as a reference to pre-conceived entities and relations (Eschenbach, 1999).

Adopting a systemic functional approach, Matthiessen (1995, 1998) and Halliday and Matthiessen (1999, 2004, 2014) described the transitivity system of English in terms of participant and circumstance

roles and Participant and Circumstance classes of semantic constituent. In order to treat spatial language, Bateman et al. (1995, 2010) defined the Element class of semantic constituents, which can be both a Thing and a Circumstance. This resulted in the Generalized Upper Model (GUM 3.0) and its spatial extension (GUM-space, Hois et al., 2009) [1], an ontology of semantic units following the principles of unique specification and minimum commitment.

However, being primarily designed after linguistic evidence in English and German, the question of to what extent GUM also applies for other natural languages, such as Brazilian Portuguese, remains partially unanswered. In this paper, we verify how far the model covers a Brazilian corpus of spatial actions and relations [2]. With our findings, we propose a reformulation of GUM's typology[3] for relational figures and conceptualize stability constraints for contextualization.

2 GUM Coverage

We collected a representative corpus of spatial relations and directed motions in Brazilian Portuguese from the tourist guide "Guia Essencial de Curitiba" (Essential Guide to Curitiba)[4]. Clauses and phrases were annotated with the terminology of GUM and GUM-space: the ones which were not predicted by the linguistic theory inside GUM were kept separate as a support for reviewing the model. After annotation, the instances of each class were inspected: when more specific linguistic variants were found under the same class, they were marked for the review phase.

Out of 304 instances of spatial figures (type of clause meaning) and spatial circumstances (type of phrase meaning), 288 (94.7%) were covered by GUM's terminology and 16 (5.3%) were not, including Examples 1-3:

(1) abrigava teatros e cafés
housed theaters and cafés
'[it] used to house theaters and cafés'

(2) o Palácio Avenida, sede do banco HSBC,
the Palace Avenida headquarters of-the bank HSBC
'the Avenida Palace, headquarters of HSBC,'

(3) o campanário da igreja com a bandeira do Brasil no topo
the bell-tower of-the church with the flag of-the Brazil at-the top
'the bell tower of the church with the Brazilian flag at its top'

Out of the covered instances, 51 (16.8%) were marked as underspecified, i.e. they are only spatial relations after contextualization (cf. Section 4), and 46 (15.1%) had an uncovered temporal variation, including the opposing pairs in Examples 4-5 and 6-7.

(4) ir para a praia
go to the beach
'going to the beach' (and staying there for a while)

(5) chegar até a Praça Espanha/Batel Soho
arrive until the square Espanha/Batel Soho
'to arrive at Espanha/Batel Soho Square' (no commitment to a longer stay)

(6) do outro lado da praça fica a entrada do Passeio Público
of-the other side of-the square is.STABLE the entrance of-the Passeio Público
'on the other side of the square is the entrance to Passeio Público'

[1] The ontology files can be downloaded at: http://www.ontospace.uni-bremen.de/ontology/gum.html

[2] Curitiba Corpus: https://docs.google.com/spreadsheet/ccc?key=0AjjU8ITs-OqudDE1MkZoS19IQWJ2Tks0NE5ONFhrZEE&usp=sharing

[3] https://github.com/DanielCoutoVale/UpperModel

[4] http://blogdapulp.wordpress.com/guias-de-viagem/guia-essencial-de-curitiba/

(7) em seu lugar estava a antiga Matriz
at its place was the old Matriz
'in its place used to be the old Matriz [church]'

2.1 Subjectless Clauses

It is worth noticing that the way Brazilian Portuguese anchors clauses to paragraph topics is different from that of English and German. While German and English always have a clausal subject related to the topic of the paragraph in the form of a noun-group, Brazilian Portuguese does not.

On the one hand, Brazilian Portuguese may conflate the subject with the finite process or auxiliary as in '**está** a três quadras da Alameda' (***it-is*** *three blocks away from Alameda*). This subject-finite conflation leaves a trace in the finite: for instance, if the subject were the speaker, the clause would be '**estou** a três quadras da Alameda' (***I-am*** *three blocks away from Alameda*).

On the other hand, Brazilian Portuguese also allows completely subject-less clauses such as 'são três quadras até a Alameda' (**are* *three blocks until Alameda*), whereby the thing which is three blocks away from the Alameda (the functional subject) leaves no trace in the clause structure because the finite process or auxiliary agrees with the direct complement. The semantics of subjectless clauses is not covered by GUM 3.0 and they account for 2 instances in the corpus.

3 Reviewing Spatial Relations

In order to extend GUM over uncovered phenomena, it was necessary to restructure the current typology of spatial relations. We remodelled the semantic constituents of spatial relations proposed by Bateman et al. (2010) by adding 4 new dimensions: *intensiveness*, *predication*, *version* and *stability*. In the remaining of this section we shall describe with examples from the corpus how this new structure of the ontology fits more accurately the flexibility of spatial language observed in Brazilian Portuguese.

3.1 Relational Intensiveness

Spatial relating figures may have two constitutional structures: the Intensive is composed by a process between two simple things, such as 'the frost covers the grass', and the Incidential is composed by a process, a spatial relative and a simple thing, such as 'the frost is on the grass'.

3.2 Relational Predication

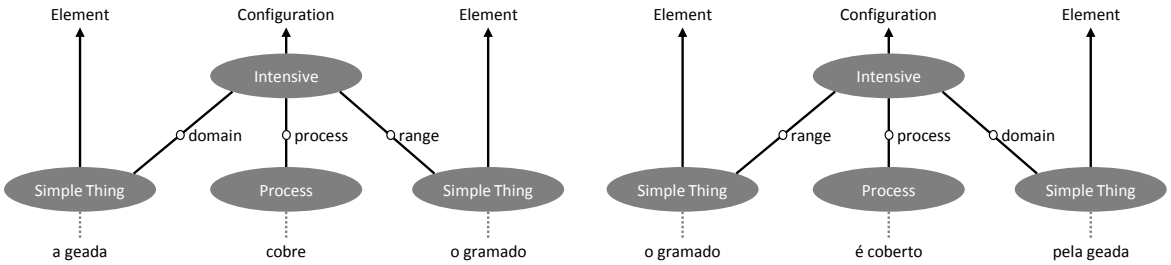
Relating figures may have different participants as the subject. A relational predicate may receive either the domain as subject (domain-receptive) or the range as subject (range-receptive).

For Intensive relating figures, Portuguese offers two predicate options: a domain-receptive predicate as in 'a geada cobre o gramado' (*the frost covers the grass*, Figure 1(a)) and a range-receptive one by reordering the constituents and inserting the auxiliary 'ser' and the case 'por' as in 'o gramado **é** coberto **pel**a geada' (*the grass* ***is*** *covered* ***by*** *the frost*, Figure 1(b)).

And for Extensive relating figures, Portuguese constructs the relational voice by varying the type of the relative constituent. The relative of the domain-receptive voice is a Circumstance composed by a relation and a relatum as in 'a geada fica em cima do gramado' (*the frost is on the grass*, cf. Figure 2(a)) and that of the range-receptive voice is a SetUp composed by a relator and a relation as in 'a grama fica com a geada em cima' (*the grass has the frost on it*, cf. Figure 2(b)).

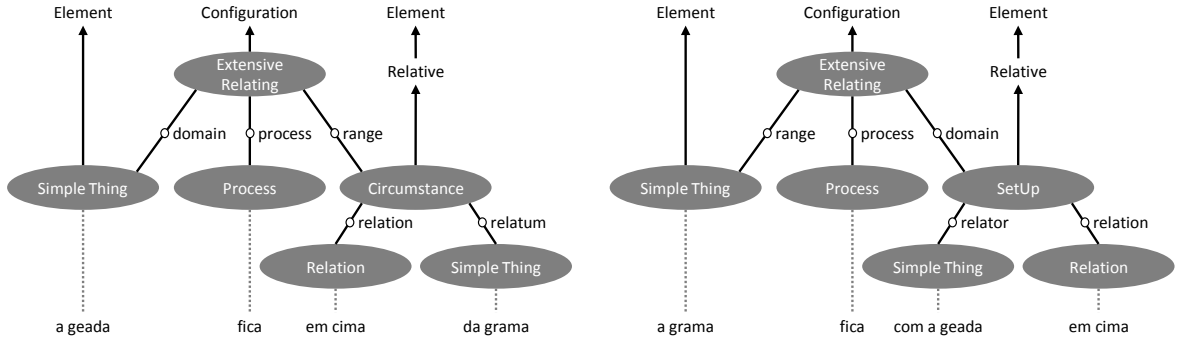
3.3 Relational Version

Known in linguistics as diathesis, another variation found in the corpus lies in the different mappings of logical roles (such as locatum or locator) to the relational roles (domain, range, *relator*, and *relatum*). We could cover all relating figures with two mappings and four diathetic roles.



(a) domain-receptive Intensive. (b) range-receptive Intensive.

Figure 1: Intensive Relating figures.



(a) domain-receptive Extensive. (b) range-receptive Extensive.

Figure 2: Extensive Relating figures.

On the one hand, the Featuring mapping makes a *carrier* out of domains and relators and a *feature* out of ranges and relata as in 'o palácio abriga a prefeitura' (*the palace houses the prefecture*, Figure 3(a)) and 'o palácio é sede da prefeitura' (*the palace is headquarters of the prefecture* 3(b)).

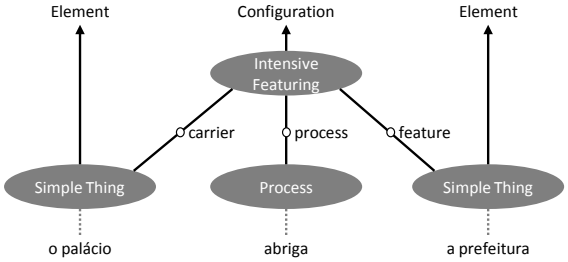
On the other hand, the Marking mapping makes a *mark* out of domains and relators and a *setting* out of ranges and relata as in 'a prefeitura habita o palácio' (*the prefecture inhabits the palace*, Figure 3(c)) and 'a prefeitura fica dentro do palácio' (*the prefecture is inside the palace*, Figure 3(d)).

For this reason, in the diathesis of spatial roles, the locatum role filled by 'a prefeitura' (*the prefecture*) specifies both *feature* and *mark* roles and the locator role filled by 'o palácio' (*the palace*) specifies both *carrier* and *setting* roles.

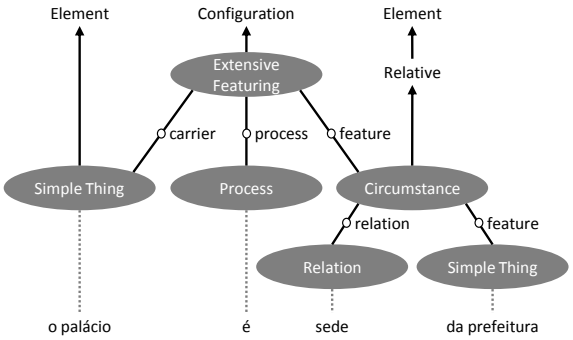
In the phrase level as qualifiers, intensiveness, voice, and version are also present. Intensive relations make domain-receptive voice with the present participle form of the process as in 'abrigando a prefeitura' (housing the prefecture) and range-receptive voice with the past participle form as in 'abrigada pelo palácio' (housed by the palace), while Incidential relations make domain-receptive voice with a *Circumstance* 'sede da prefeitura' (headquarters of the prefecture) and range-receptive with a *SetUp* with 'com' as in 'com o palácio de sede' (with the palace as headquarters), thereby leaving the process undefined, i.e. without any lexical material.

3.4 Stability of Relation and of Action Result

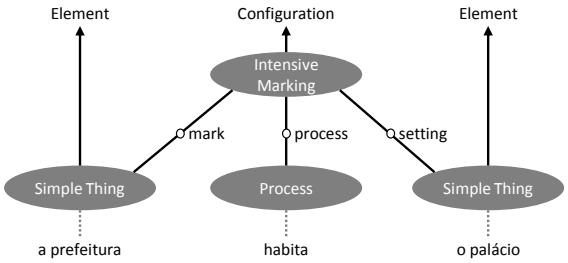
Using image schemas, (Araújo, 2008) made an analysis of spoken-language expression pairs such as 'eles estão **no** Maranhão' (*they are **at** Maranhão [but a more precise nature of this relation is not provided]*) and 'eles estão **pro** Maranhão' (*they are **at** Maranhão [but their relation to the State of Maranhão is less stable than their relation to another state]*). Such linguistic evidence shows that relational stability plays a fundamental role in specifying which kind of spatial relation is being referred to, that is, stay or residence in a Federative State. And this variation is grammaticalized as a specification of spatial



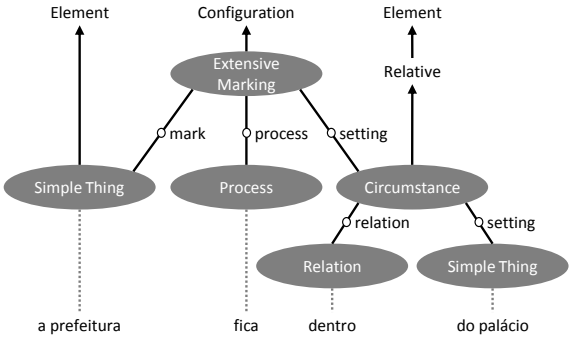
(a) Intensive and Featuring (Housing).



(b) Extensive and Featuring (Disposition).



(c) Intensive and Marking (Habitation).



(d) Extensive and Marking (Locating).

Figure 3: Featuring and Marking figures.

modality.

The same phenomenon was observable in our corpus in directed motion results. While processes such as **ir** (*go*) and **vir** (*come*) are used both for traveling and migratory movements, the spatiotemporal modality is differentiated: migratory movements take the spatial modality term **para** (*to.stable*) and traveling ones take **até** (*to*) as in 'vieram **para** a cidade' (*they migrated here to the city*) and 'vieram **até** a cidade' (*they traveled here to the city*). This means that migrants and turists share the same process of directed motion in Brazil while the kind of the changed relation between them and cities (stay or residence) is constrained by the stability contrast between the spatial modality terms **para** and **até**. This stability opposition is by no means attached exclusively to traveling and migratory movements as they also allow the contrast between 'ir **para** a praia' (*go to the beach [and stay there]*) and 'ir **até** a praça' (*go to the square [and possibly move on]*).

For Extensive spatial relations, stability was marked in our corpus by process alternation. In domain-receptive voice, the process **ficar** (*relate.stable*) specified the stability of the relation while the process **estar** (*relate*) made no stability commitments. In range-receptive voice, the process **ficar *com*** (*be-related.stable*) specifies stability and **ter** (*be-related*) makes no stability commitments. Outside our corpus, instability commitments are also to be found. The choice of the static spatial modality term **para** (*at.instable*) instead of **em** (*at*) as in 'eles estão pro Maranhão' (*they are [currently] at Maranhão*) in the spoken language corpus of Araújo (2008) specifies instability in domain-receptive voice.

4 Cultural Commitments

Our corpus shows that Brazilian Portuguese very often construes spatial relations of unspecific kind that need to be contextually understood as containment, accessibility, distance, projection or something else. In contrast, specification occurs in four other dimensions: intensiveness, voice, version, stability.

In Brazil, the type of relation is inferred from the kind of entities and the stability of the relation between them. This would justify the oppositions between 'vir para a cidade' (migrate to the city) / 'vir até a cidade (travel to the city) and 'ir para a praia' (go lay on the beach) / 'ir até a praia' (go up to the beach), in which the relational stability is used as a constraint for contextualizing the kind of relation between the person and the city or between the person and the beach.

Stability and type of entities stand for relation types not only in our corpus but also in the Brazilian culture and legislation, in which the recognition of several kinds of relations[5] are based on stability. These linguistic phenomena were not predicted by GUM 3.0, which is due to the fact that the model was created using corpora of German and English. When facing different languages, not only linguistic variation is under scrutiny, but also other underlying social phenomena.

5 Conclusion

In this paper, we have shown that Brazilian Portuguese construes space with a four-dimensional variation in intensiveness, voice, version, and stability. Supported by this linguistic evidence, we have proposed a reviewed typology of relating figure to be included in the Generalized Upper Model (GUM) and culture-specific additions such as stability to be included in a GUM extension named GUM-Brazil. With this change, we make the linguistic model fit Brazilian Portuguese flexibility more accurately and avoid supposed issues of extreme underspecification (or meaninglessness) which are barriers for applied linguistics and for the automation of linguistic analysis and synthesis in general.

References

Araújo, P. J. P. (2008). Aspectos semântico-cognitivos de usos especiais das preposições para e em na fala de comunidades quilombolas. Master's thesis, Universidade de São Paulo.

Bateman, J. A. (2010, August). Situating spatial language and the role of ontology: Issues and outlook. *Language and Linguistics Compass 4*(8), 639–664.

Bateman, J. A., R. Henschel, and F. Rinaldi (1995). The Generalized Upper Model 2.0. In *Proceedings of the ECAI94 Workshop: Comparison of Implemented Ontologies*.

Bateman, J. A., J. Hois, R. Ross, and T. Tenbrink (2010). A linguistic ontology of space for natural language processing. *Artificial Intelligence 174*(14), 1027–1071.

Eschenbach, C. (1999). Geometric structures of frames of reference and natural language semantics. *Spatial Cognition and Computation 1*(4), 329–348.

Francez, N. and M. Steedman (2006). Categorial grammar and the semantics of contextual prepositional phrases. *Linguistics and Philosophy 29*(4), 381–417.

Halliday, M. A. K. and C. M. Matthiessen (1999). *Construing experience through meaning: a language-based approach to cognition*. London/New York: Continuum.

Halliday, M. A. K. and C. M. Matthiessen (2004). *An Introduction to Functional Grammar*. New York: Oxford University Press.

Halliday, M. A. K. and C. M. Matthiessen (2014). *Halliday's introduction to functional grammar*. Routledge.

Herskovits, A. (1980). On the spatial uses of prepositions. In *Proceedings of the 18th Annual Meeting on Association for Computational Linguistics ACL 80*, pp. 1–5.

[5]Employment: Decreto-Lei n.5452/43, Art. 3o; Marriage: CR/88, Art. 226, §3o c/c CC, Art. 1.723, caput; Land ownership: CR/88, Art.183; Nationality: CR/88, Art. 12.

Herskovits, A. (1985). Semantics and pragmatics of locative expressions. *Cognitive Science 9*(3), 341–378.

Hois, J., T. Tenbrink, R. Ross, and J. A. Bateman (2009). Gum-space. The Generalized Upper Model spatial extension: A linguistically-motivated ontology for the semantics of spatial language. Technical report, Collaborative Research Center for Spatial Cognition, University of Bremen, SFB/TR8 Spatial Cognition.

Kracht, M. (2002). On the semantics of locatives. *Linguistics and Philosophy 25*(2), 157–232.

Matthiessen, C. M. (1995). *Lexicogrammatical cartography: English systems*. International Language Science.

Matthiessen, C. M. I. M. (1998). The TRANSITIVITY of space in topographic procedures. MS.

Talmy, L. (1983). How language structures space. In H. L. Pick and L. P. Acredolo (Eds.), *Spatial Orientation: Theory, Research, and Application*, pp. 225–282. New York: Plenum Press.

Zwarts, J. (2005). Prepositional aspect and the algebra of paths. *Linguistics and Philosophy 28*(6), 739–779.