SciELO Preprints

Publication status: This preprint has not been published elsewhere.

The system of identification in Brazilian Portuguese

Giacomo Figueredo

https://doi.org/10.1590/SciELOPreprints.18897

Submitted on: 2026-11-12
Posted on: 2026-11-12 (version 1)
(YYYY-MM-DD)

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

The system of IDENTIFICATION in Brazilian Portuguese

Giacomo Figueredo
Universidade Federal de Ouro Preto, Brazil
giacomo.figueredo@ufop.edu.br
https://orcid.org/0000-0002-0967-244X

Abstract: We follow systemic principles for language description – axial reasoning, text-based evidence and trinocular perspective – to describe IDENTIFICATION in Brazilian Portuguese (BP). Through agnation we detail the criteria that classify the available choices and the identification structure they can generate. A central methodological argument is that this structure will be modelled synoptically. Three systems are proposed for BP: MENTION TYPE, CO-REFERENCE, and ADJACENCY. The function structures realising features of these systems are modelled as identity graphs – mentions (nodes), identities (an edge meeting co-referential mentions) and adjacency edges. The description is demonstrated against a fully annotated BP corpus tested across genre families. Because co-reference is an equivalence relation, each co-referential mention has one identity, independently of any particular text. Several inherited categories, including retrieval relations defined by serial or temporal order, are shown to require no mechanism beyond this description.

Keywords: co-reference; IDENTIFICATION; discourse semantics structure; identity graph; Brazilian Portuguese

1 Introduction

Over the years, advances in Systemic Functional Linguistics (SFL) description and theory have recontextualised discourse semantics to capture the nature of discourse more precisely. To begin with, discourse semantic systems are realised by covariate structures (Lemke 1985) – distinct from the multivariate and univariate structures typical of grammar. Structures are of indefinite extent, and no compositional hierarchy has been found among them (Martin 2015). Rank, accordingly, is not an appropriate point of origin; the origin for discourse semantic systems is stratum.

In addition, systemic dimensions today are correspondingly better understood. They can be separated into (i) resources (axis); (ii) hierarchical organisation and bundling (rank, strata, metafunction); (iii) dynamics and behaviour (instantiation) and (iv) socialisation and allocation (individuation) (cf. Martin 2026; Doran in press; Hao 2026). As a result, discourse semantics can be described in its own terms – as a stratum of organisation in reference to axis – leaving issues of dynamics (e.g. text unfolding, flow of discourse) and socialisation (e.g. extrinsic functionality, institutions, culture) to be addressed elsewhere.

Notable work has been done on individual systems (cf. Martin and White 2005 on APPRAISAL; Martin and Dreyfus 2015 on NEGOTIATION; Hao 2020 on IDEATION). Recent work has called for new approaches to IDENTIFICATION. Quiroz (2026) shows through reference in Spanish the integrated, text-based accounts now possible for languages beyond English. She goes on to state that such accounts can advance SFL theoretically and methodologically, but require reconsidering key ideas, including how inter-stratal relations between contextual parameters and linguistic patterns are modelled across systems of choice, in explicit, paradigmatically oriented description.

As an example, we can point at IDENTIFICATION, for which the system-structure cycles familiar from grammar remain incomplete for the discourse stratum. A related problem is how discourse semantic structures are modelled. As descriptions increasingly feed into multilingual analysis and computational processing (cf. Novák et al. 2026), a representation that cannot be annotated, compared across languages or processed at scale is limited.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

In this paper, we describe the discourse semantic system of IDENTIFICATION in Brazilian Portuguese (BP) and its realising structure. Our aim is to show that discourse semantic structure can be modelled as part of the system-structure cycle, within stratification, rather than as a property of individual unfolding texts. From this, four questions follow.

(1) If covariate relations differ from grammatical structures (indefinite extent and no compositional hierarchy), what unit can serve as entry condition for a system and what realisation statement can account for the structure it generates? (2) If dynamics is set aside from the description of the stratum itself, how do we treat the apparently 'travelling' nature of text – its unfolding across space and time – without bringing in instantiation into the structure? And (3) if society, institutions and culture are likewise set aside, how can we account for IDENTIFICATION as a discourse semantic system in its own right, and what are the implications for its system-structure cycles? (4) How can a description be more useful to the analyst annotating discourse data, often at scale and across languages?

The ***descriptive parameters*** below are working hypotheses, tested as the system-structure cycles for IDENTIFICATION in BP develop.

– (i) *Entry condition*. The entry condition for IDENTIFICATION is the stratum, and the structure generated spans the whole text because its members are of indefinite extent and are not hierarchically composed. A clause ends once its function slots are filled; a text (the basic unit of discourse semantics) is not resolved by axis choices, but rather is bound by genre. The resulting structure is perhaps better modelled as a graph than as a constituency tree.

– (ii) *Relation categories*. The retrieval relations traditionally used for participant identification are mostly defined by serial order and described as a resource for the unfolding text (cf. Gleason 1968). Treating 'text unfolding' as a systemic property inevitably imports instantiation into a system network – in categories (e.g. anaphora and presenting), reactances (e.g. first mention, relevant) and basic notions (continuity, discontinuity). A systemic description (axis) for discourse semantics (stratum) should state what is configurationally possible, rather than when anything happens. We follow Martin's (1992) realisation statements for participant identification, which, synoptic by design, treat 'text-as-system' (structure) as a different object from 'text-as-process' (chains).

– (iii) *The role of context in the description*. Society, institutions, culture, cognition and the 'outside the text' are often reached for as motivation or consequences of discourse semantic patterns of identification (cf. Martin and Rose 2007; Fontaine, Jones and Schönthal 2023). As these belong to individuation, we approach discourse semantics 'from above' in relation to genre patterns, specifically how genre is realised in discourse.

– (iv) *People, things, places, and activities*. Identification makes 'identity and co-reference' meanings by 'tracking' participants – people, things, places, and activities – as they are 'introduced and then referred to subsequently in a text'. Gleason (1968) had already found that participant identification is a property of the whole text – as opposed to a 'tie' between two items or to a referent retrieved from elsewhere (the listener's mind or the world outside the text). Describing IDENTIFICATION in BP first requires defining what counts as a *mention* and stating the mechanisms for 'tracking', including multi-participant tracking (cf. (dis)continuity).

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

2 Towards a synoptic description of IDENTIFICATION

We will start this section by making a comparison with grammar and use the go-to mood structure of English as an example.

Figure 1: Example of mood system and structure in English

	Probably	this man	was	feeding	the tiger	wasn't he?
mood features	[major: indicative: declarative / modality: modalised: probability / tagged]					
mood functions	Modal. Adj.	Subject	Finite	Predicator	Complement	Moodtag

The clause in Figure 1 is described as [major], realised by inserting a Predicator; then [indicative], inserting Subject and Finite; then [declarative], sequencing them as Subject^Finite; then [tagged], adding a Moodtag; and [modalised], inserting a Modal Adjunct. This description is accurate, but the way we told it – as a sequence of 'x then y' – is misleading.

A system-structure cycle does not happen in several stages one after the other. The choices and structural function for the clause in Figure 1 did not happen one after another, at t_1: choose [major] and insert Predicator. At t_2: choose [indicative] and insert Subject and at t_3: choose Finite and so on, with the clause complete only when all steps had passed. If anything, all of it happens at once at t_0: the system network states what can be selected and the realisation states what structure results instantaneously.

This is easier to see in grammar, where an analysis table fits across a page. A whole text, especially a long or non-written one, is harder to visualise; this is compounded by the fact that actual instances of text unfold in real time as we experience them, which is perhaps why we intuitively give texts a more temporal, sequential treatment and describe them as unfolding, flowing, developing, resonating; looking back or forward; forming strings, chains, waves and ties between what came before and what will come next.

This synoptic understanding is grounded in Martin (1992), who reaches the same point. Having laid out anaphoric, cataphoric, exophoric and homophoric retrieval for participant identification, he notes that these categories raise 'a dynamic as opposed to a synoptic perspective on participant identification' (p. 126).

Adding to this issue is a recurrent terminology problem in which 'text' can mean more than one thing. A system network states a set of potential agnating features and the implications of choice in structure. This is a synoptic stance. Dynamics, differently, is accounted for in a different dimension (instantiation) as it describes how each next choice unfolds, moment by moment in an actual instance. Although both are needed if our aim is to have a comprehensive account of discourse as a phenomenon, they are not the same.

A discourse structure is not the same as an instance. A description of the stratum of discourse semantics grounded on axial reasoning must be synoptic by definition, regardless of how much text is generated. The system-structure cycle generating a mood structure (grammar) and the system-structure cycle generating an identification structure (discourse semantics) are of the same kind, regardless of whether the entry condition is a clause (thus generating a clause structure) or a stratum (thus generating a text-as-unit-of-discourse structure). By definition, neither the clause nor the text can ever generate an instance (Martin, 2008).

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

IDENTIFICATION features such as [mention] and [adjacent] are given exactly as MOOD features such as [tagged] and [modalised] – all at once, since the system network is not bound by time. The identification structure is thus complete instantaneously, rather than accumulating point by point, tie by tie and chain link by chain link as an artefact of order in listening or reading.

2.1 Motivation and descriptive parameters

To show how the synoptic approach and the descriptive parameters of Section 1 are motivated, we start with an English text, the familiar recount *At the zoo* that opens Martin's (1992: 93) chapter on IDENTIFICATION.

TEXT 1
At the zoo
One day I went to the zoo and I saw Rhinocerous I moved to a Hippopotamus I touched him and he is hand and he is big and so I went on and I saw the tiger and this man was feeding him it was eating it up Mum tod me mv on and next came then a gorilla. I had a baby gorilla. My mum tod me to move on. I saw a watch. It was 5 ock.

Following Figure 1, we place Text 1 on a roughly 9x9 grid, highlighting the words of its 35 identification-bearing mentions, a rough word-by-word approximation (cf. Martin 1992) in the generic terminology of *people, things, places, and activities*.

Figure 2: A word grid for Text 1

One	day	**I**	went	to	**the**	**zoo**	and	**I**
saw	**Rhinocerous**	**I**	moved	to	**a**	**Hippopotamus**	**I**	touched
him	and	**he**	is	hand	and	**he**	is	big
and	so	**I**	went	on	and	**I**	saw	**the**
tiger	and	**this**	**man**	was	feeding	**him**	**it**	was
eating	**it**	up	**Mum**	tod	**me**	mv	on	and
next	came	then	**a**	**gorilla**	**I**	had	**a**	**baby**
gorilla	**My**	**mum**	tod	**me**	to	move	on	**I**
saw	**a**	**watch**	It	was	5	ock		

Next, each of the 35 bold words in Figure 2 becomes a point connected by lines to other points – either within an identity (roughly, the points belonging to the same person, thing, place or activity; for example 'Hippopotamus' to 'him' to 'he') or between identities (roughly, points belonging to different people, things, places or activities, yet connected; for example 'I' to 'tiger'). Solid lines represent identities and dotted lines links between adjacent identities.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Figure 3: Connecting identity points in Text 1 grid

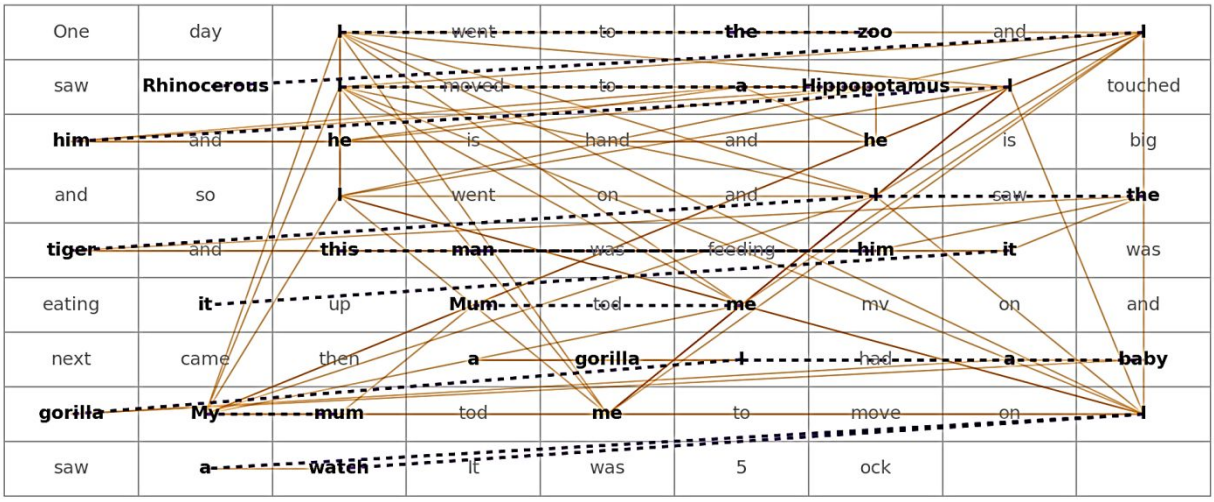


Figure 3 is revealing. It shows the whole of Text 1 as a bare instantaneous display of relations; nothing in it says 'unfolding, developing, flowing'. A line between two points states no order of occurrence. Every line is there simultaneously, as every choice in the mood structure was. Identification is thus about sameness, not order: two points are about the same person, thing, place, and activity or they are not, wherever they 'showed up' in the text.

By contrast, interpreting the lines in sequence would introduce an external temporal ordering to the relation. Any points about the same person, thing, place or activity could swap places with any others, and nothing about what is identified and linked would change. In Text 1, the ten points for the writer have one identity, all at once, as do the five for the hippopotamus, the four for the tiger, the three for the gorilla and so on.

Just as text-as-discourse-unit differs from text-as-instance, the identity structure – a functional configuration realising features of the system of IDENTIFICATION, a property of the stratum independent of any instance – differs from an identity chain, which, by contrast, is a property of a particular instance depicting successive choices dynamically. The sequential position of the points (the 'words in a text', as in each cell of Figure 3) is instantiational, not part of the structure.

Part of what the text does is distinction, not sameness though. In 'I saw the tiger', 'I moved to a Hippopotamus' and 'I had a baby gorilla', the writer is in contact with the animals without being identical to any of them, nor are the animals identical to one another. This second kind of adjacent link records that two mentions are not co-referential but still take part in the same structure.

2.2 An analogy with rhymes

Before we move on to the Methodology, perhaps it is useful to bring in an analogy to approximate the notion of *mention*, since most of the synoptic description hinges on it. The analogy is with rhymes. Text 2 is a traditional nursery rhyme analysed in Halliday and Greaves (2008), with lines numbered and rhyming syllables in bold. Three pairs of words

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

rhyme: *Muffet* /ˈm**ʌfɪt**/ (line 1) and *tuffet* /ˈt**ʌfɪt**/ (line 2); *whey* /**eɪ**/ (line 3) and *away* /əˈ**weɪ**/ (line 6); and *spider* /ˈsp**aɪdə**/ (line 4) and *beside her* /bɪˈs**aɪd ə**/ (line 5).

TEXT 2
1 Little Miss M**uffet**
2 Sat on a t**uffet**,
3 Eating her curds and wh**ey**;
4 Along came a sp**ider**,
5 Who sat down bes**ide her**
6 And frightened Miss Muffet aw**ay**.

A rhyme needs at least two words, like *whey* and *away*. *Curds* (line 3), by contrast, has nothing to rhyme with in Text 2, so it takes part in no rhyme. When there are more than two words, as there would be if a further line ended in /eɪ/, the new word joins the rhyme of *whey* and *away*. It does not start a separate rhyme with whey, apart from the rhyme with away; instead, the rhyme **accretes**. The three words 'rhyme the same rhyme', standing in a single accretive relation. A further word ending in /eɪ/ does not add anything to the rhyme; it does not expand it either. The rhyme simply accretes, remaining the very rhyme it already was.

Furthermore, the words in a rhyme are interchangeable. *Astray* (/əˈstreɪ/) can take the place of *away* in line 6 and the rhyme with *whey* stays as it was. *Chair* (/tʃeə/), on the other hand, cannot take the place of *tuffet* in line 2, *Sat on a chair*, because it does not share /ʌfɪt/ with *Muffet*. What decides whether a word can stand in is the rhyming sound it shares with the others. Neither word is prior to, or the source of, the other. *Muffet* rhymes with *tuffet* exactly as *tuffet* rhymes with *Muffet*.

For this reason, there is no rhyme-thing in Text 2 overseeing the words that rhyme. We can say that *Muffet* rhymes with *tuffet*, but we cannot point to the rhyme without pointing to those two words. A rhyme is a relation between words that rhyme. *Whey* (line 3) and *away* (line 6) are related, but not compositionally – i.e., the rhyme is not built up out of the words, nor the words out of their rhyme – and not hierarchically, with the rhyme above and the words below.

Text 2 can also be looked at discursively as the meanings of 'identity and co-reference'.

A mention is a point at which a text is about a person, thing, place or activity. Text 2 has *Little Miss Muffet* (line 1), *a tuffet* (line 2), *curds* and *whey* (line 3), *a spider* (line 4), *who* and *her* (line 5), and *Miss Muffet* (line 6). Three of these, *Little Miss Muffet*, *her* and *Miss Muffet*, are about the girl, and two, *a spider* and *who*, are about the spider. Mentions that are about the same person, thing, place or activity are co-referential, and the relation among them is an identity.

Identity stands to mentions as rhyme stands to same-sound words. Like a rhyme, an identity needs at least two mentions. *A tuffet* (line 2) has no other mention in Text 2, so it has no identity. With more than two mentions there is still a single relation – the three mentions of the girl form one identity and not three pairs. Mentions are interchangeable. *Her* can take the place of *Miss Muffet* in line 6 and the identity stays as it was, whereas substituting *Humpty Dumpty* for *Little Miss Muffet* in line 1 breaks it, since *Miss Muffet* in line 6 would then be about a different person.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

No mention is prior to, or the source of the others, and the mentions of an identity are related without composition or hierarchy. No single mention is the identity of *Miss Muffet*, and nothing in the text stands for it. We therefore speak of co-referential mentions, and not of a referent: nothing stands behind or above the mentions. An identity needs nothing beyond a relation between mentions, not even an explanation in terms of the reader's mental representations, of actual people or social life outside the text, or of something tracked or updated as the text proceeds.

Finally, mentions may be related without being co-referential. *Little Miss Muffet* and *a tuffet* are not co-referential but adjacent in *Little Miss Muffet sat on a tuffet* (lines 1 and 2). So are *who* and *her* in *who sat down beside her* (line 5). *Tuffet* and *a spider* are not adjacent, since they are not related in Text 2 at all.

3 Methodology

3.1 Corpus evidence

The data come from CALIBRA, the Catalogue of Language in Brazil (cf. Figueredo 2020): a BP corpus of roughly 1 million tokens, 803 texts, balanced across genres and socio-semiotic processes of spontaneous language use, including spoken/written, monologic/dialogic, mono/multi-semiotic modes (Figueredo 2014), with texts in the engaging, informing and arguing genre families (Rose 2019). Roughly 5% of CALIBRA, 44 texts balanced proportionally across genre family and mode (Table 1), was used for the present description. Identity and co-reference meanings were annotated by hand in this sample, alongside the development of the descriptive methodology.

The sample was already annotated for genre stages, modes, and nominal group, verbal group and clause systems (cf. Figueredo and Freitas 2026), giving direct access to most of the grammatical evidence needed for identification 'from below' and genre evidence 'from above'.

Table 1: Corpus profile

		written monolog.		written dialog.		spoken monolog.		spoken dialog.		**total**	
		texts	tokens	texts	tokens	texts	tokens	texts	tokens	**texts**	**tokens**
engaging		4	1,545	4	2,550	2	1,715	1	333	**11**	**6,143**
informing	explaining	4	1,329	4	2,455	2	707	1	1,346	**11**	**5,837**
	directing	4	1,514	4	1,449	2	1,983	1	345	**11**	**5,291**
arguing		4	1,034	4	2,943	2	674	1	320	**11**	**4,971**
total		**16**	**5,422**	**16**	**9,397**	**8**	**5,079**	**4**	**2,344**	**44**	**22,242**

3.2 Descriptive method

IDENTIFICATION is a system of discourse semantics whose domain is the text as a whole; unlike the nominal group or the clause, its entry condition is not a unit of a given rank. Methodologically, this requires ways of treating data (corpus compilation, annotation, representation) and of carrying out the description (from above, trinocular perspective, axial reasoning) so as to test our working hypotheses, checking how the description holds up against the data.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Each descriptive parameter (Section 1) is tested at a specific point. (i) Entry condition and compositional hierarchy are tested formally, by the proof that co-reference, as an equivalence relation, gives each co-referential mention one identity regardless of text. (ii) The synoptic relation between categories is tested by the graph notation: if identity and adjacency can be stated without reference to serial order or time, the parameter holds. (iii) Stratification-bound context is tested by comparing graphs of different genres under a constant annotation procedure, to see whether connectivity follows genre. (iv) The redefinition of 'people, things, places, and activities' is tested by whether MENTION TYPE, CO-REFERENCE and ADJACENCY account for cases the generic notions leave open, in particular identities that merge or branch (cf. discontinuity).

The discovery of categories in a description must be motivated by evidence and justified empirically. For IDENTIFICATION, specifically, we begin by enquiring what 'identity and co-reference' means and how we can identify them; and what this notional sense of 'sameness' (cf. referent, chain, identity) means. We adopt a trinocular perspective (Halliday 2005): relating discourse-semantic choices to genre 'from above'; describing agnate systems and their implications for structure through the identity graph 'from roundabout'; and, 'from below', noting how these choices pre-select grammatical functions (cf. Gleason 1965).

Our primary descriptive procedure is *commutation* (Hjelmslev 1961; Catford 1965; cf. Martin 1992). Commutation introduces a systematic change to an item and checks whether a change follows elsewhere. A monadic, paradigmatic commutation substitutes one item for another candidate in the same paradigm, establishing whether it is a mention at all and so forming the basis for observing choices in MENTION TYPE. A dyadic, syntagmatic commutation changes one mention and observes whether another mention co-referential with it changes, establishing the basis for co-reference.

Gleason (1968) already treats participant identification as irreducible to grammar, and applies a commutation test to separate categories, namely the descriptive, identifying and affective functions of an adjective within participant identification. Gleason's model builds participant identification cumulatively through the unfolding of the text, an Event-Line generated 'by continuous accretion at the end' (1968: 55). The present description takes up his procedure and applies it to a synoptic object.

Commutation also settles a prior methodological ambiguity: whether what we observe for identification is (a) discourse proper or (b) grammatical pre-selection. For example, are *hippopotamus* and *he* in Text 1 a discourse-semantic identity we observe directly, or a noun and a pronoun operating as Thing in the nominal group, pre-selected in turn by a discourse system generating the identity meaning? Axial reasoning resolves this by working from structure back to system – abstracting the system from agnation among functions – rather than assuming the category.

Some evidence for commutation is overt (an explicit pronoun or repeated name); some is cryptotypical (Whorf 1945), recoverable only by probing whether a co-referring mention lies within the same text or elsewhere. Further evidence is available (i) covertly, through the structural commutation extending from item to identity – substituting one established identity for another across a stretch of discourse and observing what changes in the resulting graph; (ii) overtly, through clause-rank functions associated with participants (people, things, places), realised as Participant and Circumstance; and (iii) through the grammar below – the

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

nominal group and its functions of Deictic, Thing, Numerative, Epithet, Classifier and Qualifier.

This descriptive procedure resolves into dyadic commutation for identity, since substituting a mention about a different person, thing, place or activity makes the two non-co-referential by definition. Going back to Text 1, take *tiger*, *him* and *it* in 'I saw the tiger', 'this man was feeding him' and 'it was eating it up'. If we replace one of these mentions with another mention about the tiger, they remain co-referential, simply because 'identity' cannot exist otherwise.

For mentions that are not co-referential, such as *I* and *tiger* in 'I saw the tiger' (linked by a dotted adjacency edge in Figure 3), commutation shows a different phenomenon. Let us commute *tiger* and *hippopotamus*:

'I saw the tiger' ←commutation→'I saw the hippopotamus'.

The link between the two identities, that of the writer and that of the hippopotamus, still holds because Text 1 establishes it elsewhere, as in '*I* moved to *a hippopotamus I* touched *him'*. But if we replace *tiger* with, say, *jaguar*, the link breaks because nothing in Text 1 brings *I* and *jaguar* together – there is no *jaguar* in Text 1 at all.

Text 1 is a recount (Orientation ^ Record of events) of a child walking around the zoo seeing animals. Hippopotamus is there as a recorded event. Jaguar is not. 'From above', this is why the 'I-identity' (I, I, me…) and the 'hippopotamus-identity' (hippopotamus, him) are linked. In discourse semantics, the meanings of identification make them adjacent because the zoo recount takes the writer past a hippopotamus.

3.3 Graph notation

The annotation below uses a graph notation borrowed from mathematics (Harary 1969; Bondy and Murty 2008) and inspired by Gleason (1968), who already diagrams discourse structure this way, joining a Participant to an Action by a line labelled for the Role between them. We differ from Gleason in two respects. First, our graph connects mentions to one another directly, by co-reference and by adjacency. Secondly, we assume no dynamic process of joining items, unlike Gleason.

A graph's basic element is a *node*, a single point standing for one mention in a text. A mention is a point at which the text is about a person, thing, place or activity given by commutation. It is not a piece of the text or an object it contains – like a rhyme, it is a relation.

Figure 4: A single node: one mention, 'the tiger'.

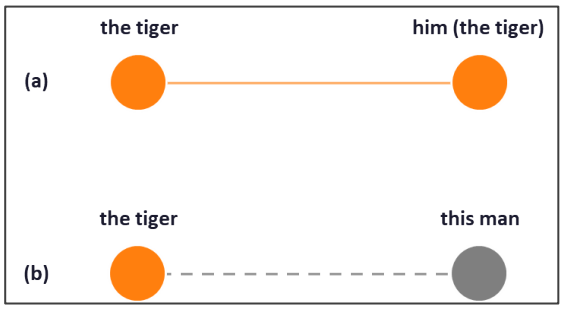


Edges connect mentions, and a graph has two kinds of edge. An identity is an edge that meets two or more co-referential mentions, at which the text is about the same person, thing, place

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

or activity. An identity is not a container for mentions, nor is it something to be 'retrieved outside discourse, in the culture, the mind, the real world'. M(i) denotes the set of mentions 'M' that the identity 'i' meets. An adjacency edge meets two mentions that are not co-referential, because one element of structure involves both.

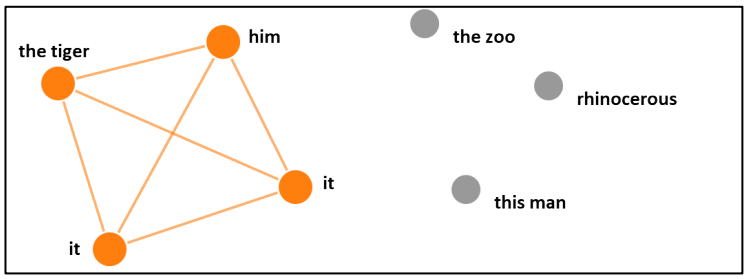
Figure 5: (a) An identity edge between 'the tiger' and 'him'; (b) an adjacency edge between non-co-referential mentions.



No edge ever means 'these two mentions are next to each other' or 'one comes before the other'; an identity edge means that two or more mentions are co-referential. 'Same' is a technical operation of commutation complemented by trinocularity and axial reasoning.

In the graph, co-reference is an equivalence relation: if mentions 'm_1' and 'm_2' are co-referential, and so are 'm_2' and 'm_3', then 'm_1' and 'm_3' must be too. Several co-referential mentions have one identity given at once, not a path of separate edges strung one after another like a chain. Drawn, the lines show a fully connected cluster (a *clique* in graph theory), but the identity is one element, not those individual lines linking nodes two by two nor a further layer added on top of those mentions. A graph built this way – identities, plus singletons (nodes not connected by any edges) – is an *equivalence graph*, the only possible shape for its identities.

Figure 6: An equivalence graph in orange; singletons in grey.

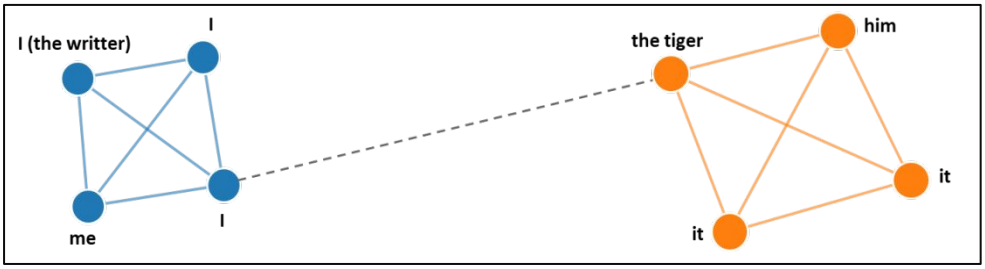


A node can carry further information, grammatical or contextual (e.g. nominal group pre-selection, clause grammar function, genre stage – whatever serves the research). None of this affects the shape of the graph but is useful for observing correlated patterns.

A graph can show more than one identity at once. The adjacency between them is given by dyad commutation and is represented by a dotted line. In Text 1, 'I' (the writer) and 'tiger' have an identity each, and in 'I saw the tiger' one manifestation involves both. The two identities are linked by an adjacency edge.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Figure 7: Two identities and an adjacency edge between them.



In traditional graph design, a layout algorithm places nodes wherever the picture reads most easily, and two people drawing the same graph may place a node differently without disagreeing about anything. In our description, where a node is on the page carries meaning. Every node has a coordinate given by the number of relations it establishes with other nodes: the more connected a node, the more it attracts other nodes.

The result is a force-driven graph layout (Fruchterman and Reingold 1991). Every pair of nodes repels with a force 'f_r', a function of the distance '(d)', equal to the ideal distance 'k' squared over the distance 'd'.

$$f_r(d) = \frac{k^2}{d} \tag{1}$$

The ideal distance 'k' between nodes 'n' is given by the equation:

$$k = \frac{1}{\sqrt{n}} \tag{2}$$

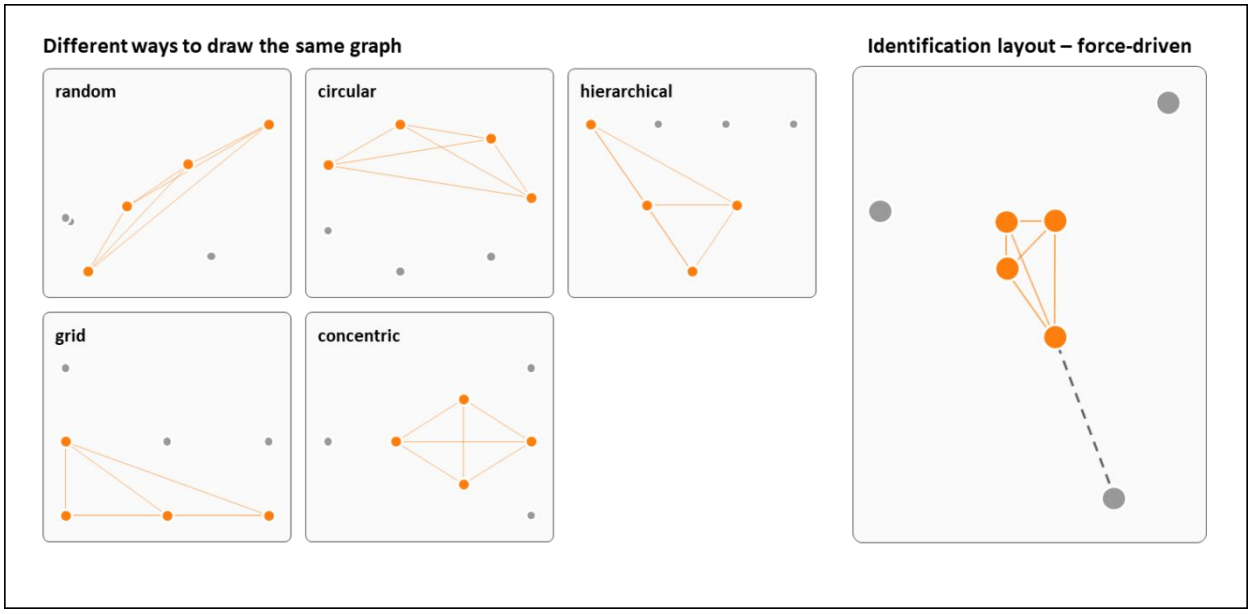
Every edge, in turn, attracts the nodes it meets with a force of attraction 'f_a', a function of the distance, equal to the weight of the edge 'w' times the distance squared over the ideal distance 'k'.

$$f_a(d) = \frac{wd^2}{k} \tag{3}$$

The weight 'w' is set at 3.0 for an identity edge and 0.55 for an adjacency edge. The nodes start from a fixed random placement (seed 13). In each of 50 iterations, every node moves in the direction of the sum of the forces on it, by a smaller step each time, until the nodes settle where pushes and pulls balance. The calculation uses the *spring_layout* function of the Python library networkx (Hagberg, Schult and Swart 2008), so annotators with the same weights and seed obtain the same coordinates.

Figure 8: Graph layouts and the force-driven identification layout used in this paper.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897



4 Identification in Brazilian Portuguese

Let us begin by introducing a text sample. It stands alone as a complete text.

TEXT 3 [PB_CUMPR_01]

Bom dia!!	Good morning!!

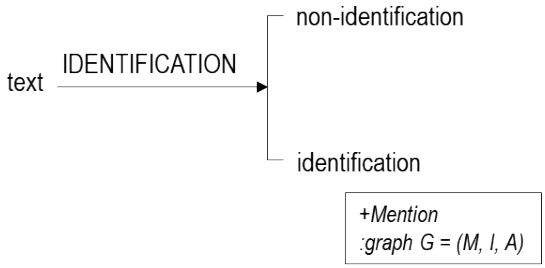
Text 3 is a greeting, which 'from above' enacts and renews contact between people. Unlike genres developed in stages, such as Orientation in a narrative, or Topic in a report, it introduces nothing.

As far as identification is concerned, a text may have nothing to identify (that is, no mention and no identity). Thus, the limiting case is a text with no identification structure, like Text 3. In this sense, it could be seen, as it were, as a 'minor text'. Looked at 'from above', this minor text does not realise any staging at all, but a single, unstaged element. If we compare it with a service encounter, where there is also a greeting, we can see the difference. In a service encounter, the greeting is not a whole text, but one stage in a larger staged genre (Greeting ^ Service ^ Closing) (cf. Martin 1985).

This is the first agnation in the system: a text either selects [identification] or [non-identification]. Text 3 selects [non-identification]. If a text, differently, selects [identification], then it must insert an identification structure – the function Mention, which preselects a class of graph (+Mention; Mention: graph).

Figure 9: First agnation in IDENTIFICATION.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897



Systemically speaking, a graph is a class of discourse semantics, as *verbal group*, *noun* and *material clause* are classes of grammar. A discourse semantic function, such as Mention, preselects items of this class, as an element of structure in the grammar preselects classes of word, group, and so on. And, most importantly, as Text 3 shows, the entry condition for IDENTIFICATION is the whole text, since there is no compositional hierarchy, no rank and the structure is co-variate.

Graph theory names most of the types of graph but does not classify them, certainly not by what a graph does in a text. The names we use here are borrowed from graph theory, and the classes are defined as classes of discourse semantics. As introduced in Section 3 above, a graph is a set of nodes and a set of edges, with each edge joining two nodes. Discursively, a mention is a node. A full node (represented by ●) is one that edges can join. An empty node (represented by ○) is one that no edge can reach (Table 2).

Table 2: Classes of graph

graph class	**graph theory**	**nodes**	**edges**	**drawing**	**whole text**
null	null graph	none	none	(nothing)	none: [non-identification], inserts nothing (Text 3)
trivial	trivial graph	one	none	● or ○	(a), (b) below
discrete	discrete (edgeless) graph	n	none	○ ○ ○	e.g. Text 4, lines 2-7
complete	complete graph	n ≥ 2	every pair	●–●	identity, co-reference
simplex	simplex	n ≥ 1	every pair; one element	●, ●–●, filled triangle, tetrahedron, etc.	identity, co-reference
dyad	no standard name	two or more simplexes	edges between non-equivalent simplexes; none inside one	●–● ⋯ ●–●	adjacency

A trivial graph is a single node, and the node may be full or empty. In (a) the text is about *roupa* ‘clothing’ as a mass, an empty node because it cannot have an edge with another mention to form an identity; in (b) it is about *roupa* as an item ‘an outfit’, which can.

(a) *vend-o* ***roupa*** → one empty node ○
sell-PRS.1SG **clothes.MASS**
‘I sell clothing.’

(b) *vend-o* ***uma*** ***roupa*** → one full node ●
sell-PRS.1SG **INDF.F.SG** **clothes**
‘I sell an outfit.’

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Simplex is used in the topological sense, as the simplest figure in each dimension, not in the simplex/complex opposition of logical structure. 'n' nodes joined at once make an (n – 1) simplex. One node is a (1 – 1) = zero-simplex; two nodes and an edge, a (2 – 1) = one-simplex; three nodes and a filled triangle, a (3 – 1) = two-simplex; and so on. Structurally, a simplex is the whole shape mapped onto by one element. The term *dyad*, borrowed from structural commutation, names two or more non-equivalent nodes as a single element of structure.

4.1 MENTION TYPE, Isolate, and Incident

Let us look at Text 4 – a taxonomic report.

TEXT 4 [PB_RELTAXO_01]

Topic	1	Características dos cinco reinos dos seres vivos	Characteristics of the five kingdoms of living things
Description	2	Nutrição - Autotrófica ou heterotrófica	Nutrition – Autotrophic or heterotrophic
	3	Organização celular - Unicelulares ou pluricelulares	Cellular organisation – Unicellular or multicellular
	4	Tipologia celular - Eucariontes ou procariontes	Cell typology – Eukaryotic or prokaryotic
	5	Respiração - Aeróbia ou anaeróbia	Respiration – Aerobic or anaerobic
	6	Reprodução - Sexuada, assexuada ou por esporos	Reproduction – Sexual, asexual, or by spores
	7	Locomoção - Autônoma ou imóvel	Locomotion – Autonomous or immobile

'From above', a taxonomic report classifies things and describes the classes. Its generic structure is Topic ^ Description. The Topic in Text 4 names the classification, *Características dos cinco reinos dos seres vivos* 'Characteristics of the five kingdoms of living things'.

If we concentrate on the Description stage, in 2–7, we find a taxonomy of activities, such as *nutrição* 'nutrition' in (2), *respiração* 'respiration' in (5) and *locomoção* 'locomotion' in (7). Each states the activity first and its kinds after it, so that (5), *Respiração – Aeróbia ou anaeróbia* 'Respiration – Aerobic or anaerobic', is about *respiração aeróbia* 'aerobic respiration' and *respiração anaeróbia* 'anaerobic respiration'. The Description is about classes and sets up no particular person, place, thing or activity for identities. Text 4's stages are about something, so it selects [identification] and has a graph. Its Description stage has mentions, but *discrete* ones that can never establish edges with other mentions.

We can test this from roundabout through commutation. We can commute *respiração aeróbia* 'aerobic respiration' in (5) with *digestão intracelular* 'intracellular digestion'. And it changes nothing else. The same holds for all the others (Table 3).

Table 3: Commutation in Text 4

line	**source text**	**commuted text**	**change elsewhere**
5	*respiração aeróbia* 'aerobic respiration'	*digestão intracelular* 'intracellular digestion'	none
7	*locomoção autônoma* 'autonomous locomotion'	*deslocamento autônomo* 'autonomous displacement'	none
1	*os seres vivos* 'the living things'	*os organismos* 'the organisms'	none
7	*locomoção imóvel* 'immobile locomotion'	*respiração aeróbia* 'aerobic respiration'	none

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Commutation separates the isolated mention from mentions that could form part of the same identity as other mentions. Text 4 mentions only isolated activities, unrelated to any other by identity or co-reference. This contrast picks up the agnation of the two features of the system of MENTION TYPE – [isolate] and [incident]. Both are mentions: an isolate mention is an empty node (○) and an incident mention is a full node (●) – or a node an edge could potentially meet.

From below, (5) and (7) (Table 4) have the wording that goes with an [isolate] – a Thing that is continuous (i.e. mass) – as do the other nodes in 2–7.

Table 4: Nominal groups of two nodes in Text 4

line	**wording**	**nominal group**	**node**
5	*respiração aeróbia* 'aerobic respiration'	Thing *respiração*, Classifier *aeróbia*	empty
7	*locomoção autônoma* 'autonomous locomotion'	Thing *locomoção*, Classifier *autônoma*	empty

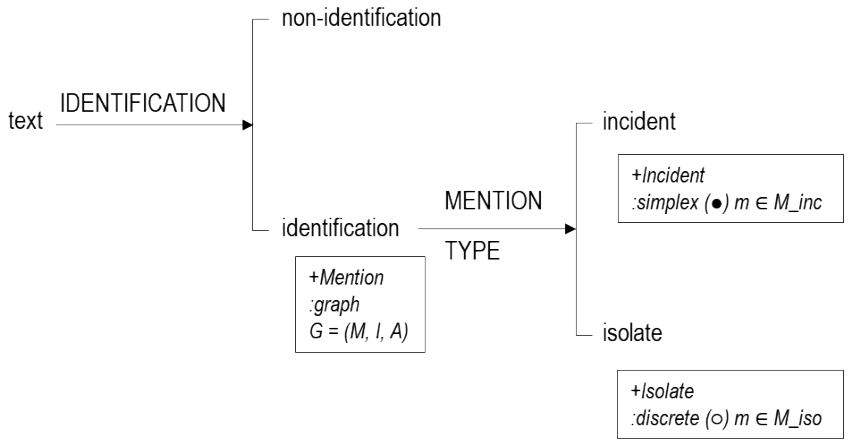
The features of MENTION TYPE are realised in discourse semantics as functions in the structure of discourse preselecting classes of graph. [identification] inserts the function Mention (+Mention), which preselects the class graph; [isolate] inserts Isolate and [incident] Incident, each preselecting a class of graph as follows:

Table 5: Features, functions and classes of graph

feature		**realisation**	**in the graph**	**notation**
[non-identification]		nothing	no graph	–
[identification]		+Mention	graph	G = (M, I, A)
	[isolate]	+Isolate: discrete	empty node ○; class discrete	$m \in M_{iso}$
	[incident]	+Incident: simplex	full node ●; class simplex	$m \in M_{inc}$

In notation, G is the graph. M is the set of all mentions, M_{inc} is the set of mentions that select [incident] and M_{iso} the set of mentions that select [isolate]. A mention selects one feature or the other, so $M = M_{inc} \cup M_{iso}$, and $M_{inc} \cap M_{iso} = \emptyset$.

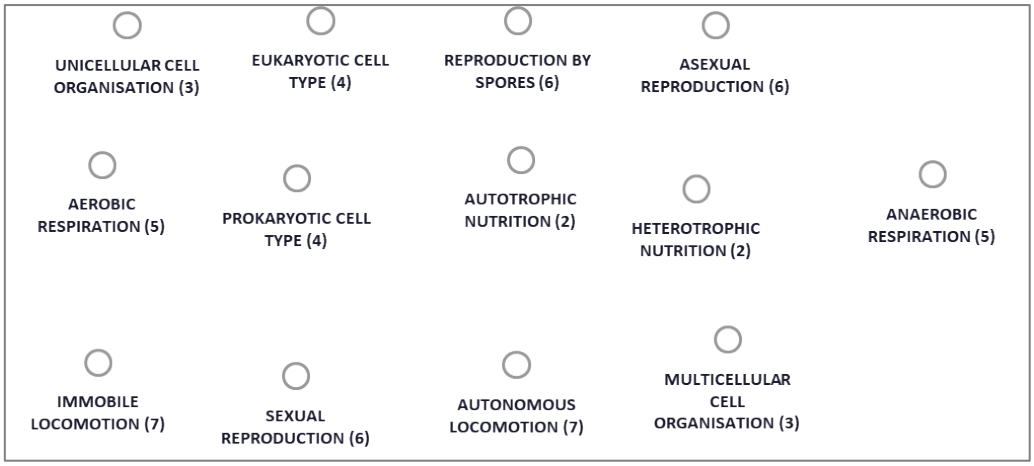
Figure 10: MENTION TYPE.



The Description stage of Text 4 has thirteen nodes. All are mentions selecting [isolate]: empty nodes with no edge joining, so 2–7 form a discrete graph (Figure 11).

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Figure 11: Identification structure of the Description stage of Text 4, modelled as a graph – discrete, empty nodes.



4.2 Co-REFERENCE and Identity

We begin this section by introducing Text 5. It is a Brazilian folk tale of the narrative genre, in which the Poor Man and the Rich Man trick each other over a piece of land and the magical treasure found on it.

TEXT 5 [PB_NARR_01]

Orientation	1	Havia dois homens, um rico e outro pobre, [[que gostavam de pregar peças um ao outro]].	There were two men, one rich and one poor, [[who liked playing tricks on each other]].
	2	Foi o compadre pobre à casa do rico	The poor man went to the rich man's house
	3	pedir um pedaço de terra	to ask for a piece of land
	4	para fazer uma roça.	to farm.
	5	O rico, para fazer peça ao outro, lhe deu a pior terra [[que tinha]].	The rich man, <<to play a trick on him>>, gave him the worst land [[he had]].
	6	Logo que o pobre teve o sim,	As soon as the poor man had the 'yes', (the rich man's say-so)
	7	foi para a casa dizer à mulher,	he went home to tell his wife,
	8	e foram ambos ver o terreno.	and the two of them went to see the land.
Complication 1	9	Chegando lá nas matas,	Getting there in the woods,
	10	o marido viu uma cumbuca de ouro,	the husband saw a pot of gold,
	11	e, <<como era em terras de compadre rico>>, o pobre não a quis levar para a casa,	and, <<since it was on the rich man's land>>, the poor man did not want to take it to the house,
	12	e foi dizer ao outro	and went to tell the other man
	13	que em suas matas havia aquela riqueza.	that there was that treasure on his land.
	14	O rico ficou logo todo agitado,	The rich man was at once all worked up,
	15	e não quis	and would not have it
	16	que o compadre trabalhasse mais nas suas terras.	that the poor man should work his land any longer.
	17	Quando o pobre se retirou,	When the poor man left,
	18	o outro largou-se com a sua mulher para as matas	the rich man set off with his wife for the woods
	19	a ver a grande riqueza.	to see the great treasure.
	20	Chegando lá,	Getting there,

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Complication 2	21	o [[que achou]] foi uma grande casa de marimbondos;	[[what he found]] was a huge wasps' nest;
	22	meteu-a numa mochila	he put it in a bag
	23	e tomou o caminho do mocambo do pobre,	and took the path to the poor man's shack,
	24	e logo que o avistou foi gritando:	and <<as soon as he caught sight of him>> he called out:
	25	'Ó compadre, fecha as portas,	'Friend, shut the doors,
	26	e deixa somente uma banda da janela aberta!'	and leave just one side of the window open!'
	27	O compadre assim fez,	The poor man did just that,
	28	e o rico, <<chegando perto da janela>>, atirou a casa de marimbondos dentro da casa do amigo,	and the rich man, <<getting up to the window>>, threw the nest of wasps into the friend's house,
	29	e gritou:	and shouted:
	30	'Fecha a janela, compadre!'	'Shut the window, friend!'
Resolution	31	Mas os marimbondos bateram no chão,	But the wasps hit the floor,
	32	transformaram-se em moedas de ouro,	turned into gold coins,
	33	e o pobre chamou a mulher e os filhos	and the poor man called his wife and children
	34	para as ajuntar.	to gather them up.
Coda	35	O ricaço gritava então:	The rich man then called out:
	36	'Ó compadre, abra a porta!'	'Friend, open the door!'
	37	Ao que o outro respondia:	To which the poor man answered:
	38	'Deixe-me,	'Leave me be,
	39	que os marimbondos estão-me matando!'	the wasps are stinging me to death!'
	40	E assim ficou o pobre rico,	And so the poor man stayed rich,
	41	e o rico ridículo.	and the rich man stayed a fool.

'From above', the generic structure of Text 5 is Orientation ^ Complication 1 ^ Complication 2 ^ Resolution ^ Coda. The purpose of a narrative is for characters to resolve a complication, so each stage mentions characters (particular people, places and things). For example, Orientation (1–8) mentions the Rich Man and the Poor Man, *um rico e outro pobre* 'a rich man and another, poor' (1), the house, *a casa do rico* 'the rich man's house' (2), and the land, *um pedaço de terra* 'a piece of land' (3).

Some characters are mentioned in more than one stage. In Orientation (1–8), the Rich Man is mentioned at *um rico* 'a rich man' (1), *do rico* 'of the rich man' (2) and *o rico* 'the rich man' (5). The Poor Man is mentioned at *outro pobre* 'another, poor' (1), *o compadre pobre* 'the poor friend' (2), *ao outro* 'to the other one' (5) and *o pobre* 'the poor man' (6). Complication 1 (9–19) mentions the Poor Man at *o marido* 'the husband' (10), *o pobre* (11, 17) and *o compadre* 'the friend' (16). It mentions the Rich Man at *compadre rico* 'rich friend' (11), *ao outro* 'to the other one' (12), *o rico* (14) and *o outro* 'the other one' (18). Complication 2 (20–30), the Resolution (31–34) and the Coda (35–41) mention them again.

From roundabout, commutation tests whether these are mentions in identification.

Table 6: Commutation in Text 5

line	**source**	**commuted**	**change elsewhere**
1	*um rico* 'a rich man'	*um irmão* 'a brother'	*o rico* (5, 14, 28), *o outro* (18) no longer same man as (1)
10	*uma cumbuca de ouro* 'a gold pot'	*uma casa de marimbondos* 'a wasps' nest'	*aquela riqueza* (13), *a grande riqueza* (19) no longer same thing as (10)

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

14	*o rico* 'the rich man'	*o pobre* 'the poor man'	no longer same man as *um rico* (1); now same as *o compadre pobre* (2)
33	*os filhos* 'the children'	*os vizinhos* 'the neighbours'	none: no other mention same as it
18	*a sua mulher* 'his wife'	*a sua cunhada* 'his sister-in-law'	none: no other mention same as it

In the first three rows something else in the text changes with the commutation: other mentions are the same character, or the same thing, as the one replaced. The mentions are of a particular character (the Rich Man) and a particular pot, and each selects [incident]: other mentions can be the same as it. In the last two rows the commutation changes none: no other mention in the text is the same as it.

This contrast picks up the agnation in a more delicate system for [incident], that of CO-REFERENCE. A delicacy of [incident] can be [co-referential], when another mention in the text is the same as it – forming an identity – or [singleton], when none is.

From below, each mention is worded in BP as a nominal group with a Thing and, generally, a Deictic (*um*, *uma*, *o*, *os* 'a, the'), an Epithet (*grande* 'huge') or a Classifier (*de ouro*, *de marimbondos* 'gold, wasp') (cf. Figueredo 2021).

Table 7: Nominal groups of mentions in Text 5

stage	**line**	**wording**	**nominal group**	**feature**
Orientation	1	*um rico* 'a rich man'	Deictic *um*, Thing *rico*	[incident]
Complication 1	10	*uma cumbuca de ouro* 'a gold pot'	Deictic *uma*, Thing *cumbuca*, Classifier *de ouro*	[incident]
Complication 2	21	*uma grande casa de marimbondos* 'a huge wasp nest'	Deictic *uma*, Epithet *grande*, Thing *casa*, Classifier *de marimbondos*	[incident]
Resolution	32	*moedas de ouro* 'gold coins'	Thing *moedas*, Classifier *de ouro*	[incident]
	33	*os filhos* 'the children'	Deictic *os*, Thing *filhos*	[incident]

Table 8: Mass and countable Thing in Texts 4 and 5

Thing	**BP grammar**	**wording**	**feature**
Continuous (mass)	Thing, no Deictic or Numerative	*respiração aeróbia* 'aerobic respiration' (Text 4); *vend-o roupa* 'I sell clothing' (Table 2, a)	[isolate]
Discrete	Plural Thing, Deictic or number	*os filhos* 'the children' (33); *uma cumbuca de ouro* 'a gold pot' (10); *vend-o uma roupa* 'I sell an outfit' (Table 2, b)	[incident]
Classifier	*de* + noun, no Deictic	*ouro* in *uma cumbuca de ouro* (10); *marimbondos* in *uma grande casa de marimbondos* (21)	no mention of its own; whole group is the mention

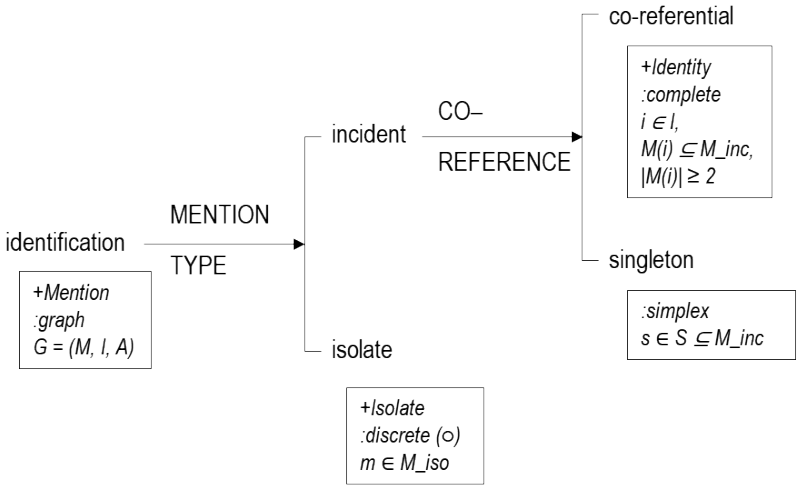
Each of these mentions selects [incident]. The feature inserts the function Incident and is realised as a simplex (Table 5). A simplex is the class of graph for an element of structure, and a single mention is the simplest, a zero-simplex. It is a full mention, which means that other mentions can be the same as it. From below, we also take note of the Deictic of some of these mentions. The feature [co-referential] inserts the function Identity (+Identity), and [singleton] inserts nothing and keeps the class simplex that +Incident preselects.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Table 9: Deictic of mentions in Text 5

line	wording	Deictic	feature
5	*o rico* 'the rich man'	*o* 'the'	[co-referential]
18	*o outro* 'the other one'	*o* 'the'	[co-referential]
13	*aquela riqueza* 'that treasure'	*aquela* 'that'	[co-referential]
19	*a grande riqueza* 'the great treasure'	*a* 'the'	[co-referential]
18	*a sua mulher* 'his wife'	*sua* 'his'	[singleton]
33	*os filhos* 'the children'	*os* 'the'	[singleton]

Figure 12: The system of CO-REFERENCE



Now we are in a position to define what 'identity' means in this paper. An Identity is a function, the element of discourse semantic structure that meets two or more mentions given by commutation. In discourse there are only mentions, and the identity is the relation among those that are the same as one another, given at once. *Um rico* 'a rich man' (1), *o rico* 'the rich man' (5, 14, 28) and *o outro* 'the other one' (18) are mentions of one identity.

The function of Identity preselects the class of graph for an equivalence relation, called 'complete'. Drawn as lines, the mentions of one identity form a complete graph. A singleton is a mention that selects [incident] that no identity meets. 's' is the set of singletons: $s \in M_{inc}$ with $s \notin M(i)$ for every $i \in I$.

In notation, I is the set of identities and $M(i) \subseteq M_{inc}$ the set of mentions that identity 'i' meets. An identity meets at least two mentions, so $|M(i)| \geq 2$. The relation '~' (read 'is co-referential with') holds between mentions met by the same identity: $m_1 \sim m_2$. It is reflexive ($m_1 \sim m_1$), symmetric (if $m_1 \sim m_2$, then $m_2 \sim m_1$) and transitive (if $m_1 \sim m_2$ and $m_2 \sim m_3$, then $m_1 \sim m_3$). It is therefore an equivalence relation on M_{inc}, and no mention belongs to the M(i) of two identities. Each feature preselects a class of graph (Table 10).

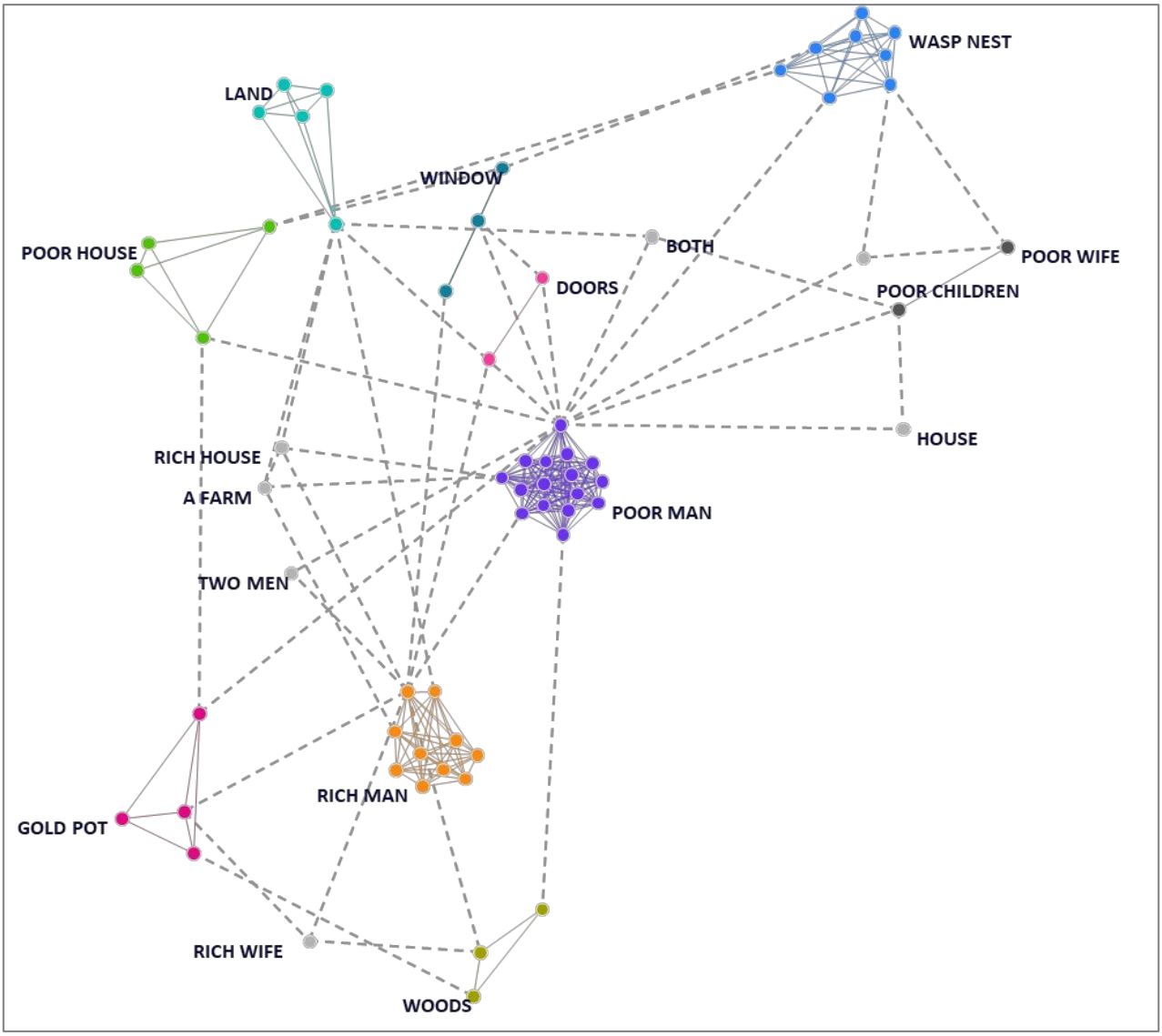
Table 10: Features, functions and classes of graph in CO-REFERENCE

feature	realisation	in the graph	notation
[co-referential]	+Identity :complete	identity meeting two or more full nodes ●, every pair joined; one element; class complete	$i \in I$, $M(i) \subseteq M_{inc}$, $M(i)\| \geq 2$
[singleton]	nothing	full node ●, no identity; class simplex, inherited from +Incident (zero-simplex)	$s \in S \subseteq M_{inc}$, $s \notin M(i)$ for every $i \in I$

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Figure 13 draws the identities of Text 5. Each coloured cluster is one identity (POOR MAN: purple; RICH MAN: orange; etc.). The more mentions an identity meets, the larger its 'identity cluster'. A grey node standing alone is a singleton, as with *a sua mulher* 'his wife' (18) and *os filhos* 'the children' (33) – they are related to the Poor Man, but they are not part of the 'Poor Man-identity cluster'. The other identities, such as GOLD POT and WASP NEST, are smaller.

Figure 13: The graph of Text 5 – the model of discourse semantics identity structure.



The characters, who resolve the complication, are the largest identities. The Poor Man is joined to more elements than the Rich Man, which is why his 'identity cluster' is drawn into the middle of the graph by the force of attraction, with the others surrounding it. Poor House, for example, sits between Poor Man and Wasp Nest, near the singletons Poor Wife and Poor Children, and far from Rich Wife. The Rich Man and the other identities are joined to fewer elements and so are pushed to the margins of the graph by the force of repulsion. Figure 13 is not merely a visual illustration but the shape taken by the discourse semantic structure that the narrative demands of IDENTIFICATION.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

4.3 ADJACENCY

Text 6 is a biographical recount of mathematician Évariste Galois.

TEXT 6 [PB_BIO_01]

Orientation	1	Évariste Galois nasceu nas proximidades de Paris, na aldeia de Bourg la-Reine,	Évariste Galois was born near Paris, in the village of Bourg la-Reine,
	2	onde seu pai era prefeito.	where his father was mayor.
Record of events	3	Aos 12 anos mostrava pouco interesse por Latim, Grego e Álgebra	At 12 he showed little interest in Latin, Greek and Algebra
	4	mas a Geometria de Legendre o fascinava.	but Legendre's Geometry fascinated him.
	5	Aos 16 anos, <<julgando-se em condições>>, procurou entrar na Escola Politécnica	At 16, <<judging himself ready>>, he tried to enter the École Polytechnique
	6	mas foi recusado por falta de preparo	but was refused for lack of preparation
	7	e isto marcou seu primeiro fracasso.	and this marked his first failure.
	8	Aos 17 anos escreveu um artigo [[onde expôs suas descobertas fundamentais]]	At 17 he wrote a paper [[in which he set out his fundamental discoveries]]
	9	entregando-o a Cauchy	handing it to Cauchy
	10	para que o apresentasse na Academia.	for him to present it at the Academy.
	11	Cauchy perdeu seu trabalho	Cauchy lost his work
	12	e com isto veio o seu segundo fracasso marcante.	and with this came his second, striking failure.
	13	Logo mais perdeu o pai [[que, devido a intrigas clericais, se suicidou]].	Soon after he lost his father, [[who, because of clerical intrigues, committed suicide]].
	14	Desiludido, Galois entrou na Escola Normal	Disillusioned, Galois entered the École Normale
	15	para preparar-se	to prepare himself
	16	a fim de ensinar,	in order to teach,
	17	sempre continuando com suas pesquisas.	always continuing with his research.
	18	Em 1830 escreveu um artigo para concurso de Matemática da Academia	In 1830 he wrote a paper for the Academy's Mathematics competition
	19	entregando-o para Fourier,	handing it to Fourier,
	20	que morreu logo depois	who died soon afterwards
	21	e o artigo foi perdido.	and the paper was lost.
	22	Com tantas frustrações Galois acabou por aderir às causas da Revolução de 1830,	With so many frustrations Galois ended up joining the causes of the Revolution of 1830,
	23	foi expulso da Escola Normal	was expelled from the École Normale
	24	e mais tarde entrou para a guarda nacional.	and later joined the national guard.
	25	Galois iniciou suas pesquisa com um trabalho de Lagrange sobre permutações de raízes,	Galois began his research following a work by Lagrange on permutations of roots,
	26	o que lhe deu condições necessárias e suficientes para concluir	which gave him the necessary and sufficient conditions to conclude
	27	que equações polinomiais são resolúveis por radicais	that polynomial equations are solvable by radicals
	28	e, <<baseado nas provas de Abel>>, descobriu	and, <<based on Abel's proofs>>, discovered
	29	que as equações algébricas irredutíveis são resolúveis por radicais somente se o grupo de permutações sobre suas raízes, assim como outros postulados sempre voltados mais para a estrutura algébrica do que para	that irreducible algebraic equations are solvable by radicals only if the group of permutations on their roots, as well as other postulates always turned more to

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

		casos específicos, dando um tratamento aritmético à Álgebra.	algebraic structure than to specific cases, giving an arithmetic treatment to Algebra.
	30	Em suas obras está implícito o conceito de "corpo"	In his works the concept of "field" is implicit
	31	que mais tarde Dedekind definira de forma explícita.	which Dedekind would later define explicitly.
	32	Na época Galois entregou a Poisson um artigo contendo sua teoria	At the time Galois handed Poisson a paper containing his theory
	33	e este o classificou de "incompreensível"	and he classified it as "incomprehensible"
	34	mas hoje o que chamamos de "Matemática Moderna" nada mais é do que as ideias de Galois [[que estão chegando até nós]].	but today what we call "Modern Mathematics" is nothing more than Galois's ideas [[that are reaching us]].
	35	Em 1832, envolvendo-se com uma mulher, em nome de um código de honra, não pode evitar um duelo.	In 1832, getting involved with a woman, in the name of a code of honour, he could not avoid a duel.
	36	Na noite anterior passou as horas rascunhando notas para a posteridade numa carta a seu amigo.	The night before he spent the hours drafting notes for posterity in a letter to his friend.
	37	Na manhã de 30 de maio encontrou seu adversário recebendo um tiro fatal.	On the morning of 30 May he met his adversary, receiving a fatal shot.
	38	Socorrido por um camponês, morreu num hospital para onde foi levado, aos 20 anos de idade.	Helped by a peasant, he died in a hospital to which he was taken, at the age of 20.

The generic structure of Text 6 is Orientation ^ Record of events. Its purpose is to record what its subject does and what happens to them. In the Orientation, Galois, Paris and the village are joined to one another in *Évariste Galois nasceu nas proximidades de Paris, na aldeia de Bourg la-Reine* 'Évariste Galois was born near Paris, in the village of Bourg la-Reine' (1), and the village and his father are joined in another, *onde seu pai era prefeito* 'where his father was mayor' (2).

In the Record of events, Galois is mentioned along with other mentions that form no identity with Galois: *Cauchy* and *o artigo* 'the paper' in *entregando-o a Cauchy* 'handing it to Cauchy' (9), *Fourier* and *o artigo* 'the paper' in *entregando-o para Fourier* 'handing it to Fourier' (19) and *Poisson* and *o artigo* 'the paper' in *Galois entregou a Poisson um artigo* 'Galois handed Poisson a paper' (32). In (21), by contrast, in *o artigo foi perdido* 'the paper was lost' the paper is mentioned on its own.

From roundabout, commutation on a dyad replaces the whole of an element, an identity or a singleton, by another and shows which element it is joined to, if any (Table 11).

Table 11: Commutation in Text 6

line	**source**	**commuted**	**change elsewhere**
5	*a Escola Politécnica* 'the École Polytechnique'	*Latim* 'Latin'	Galois identity-cluster joined to *Latim*, not *a Escola Politécnica*
11	*seu trabalho* 'his work'	*a guarda nacional* 'the national guard'	*Cauchy* identity-cluster joined to *a guarda nacional*, not *seu trabalho*
21	*o artigo* 'the paper'	*Poisson*	none: element joined to no other

In the first two rows commutation changes the dyad; in the last it changes none. This contrast picks up the agnation of the two features of the system of ADJACENCY, a delicacy of [incident]. [adjacent] is selected when the mention is joined to another element, an identity or a singleton, and [non-adjacent] when it is joined to none. *a Escola Politécnica* (5) is [adjacent] and *o artigo* 'the paper' (21) is [non-adjacent].

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

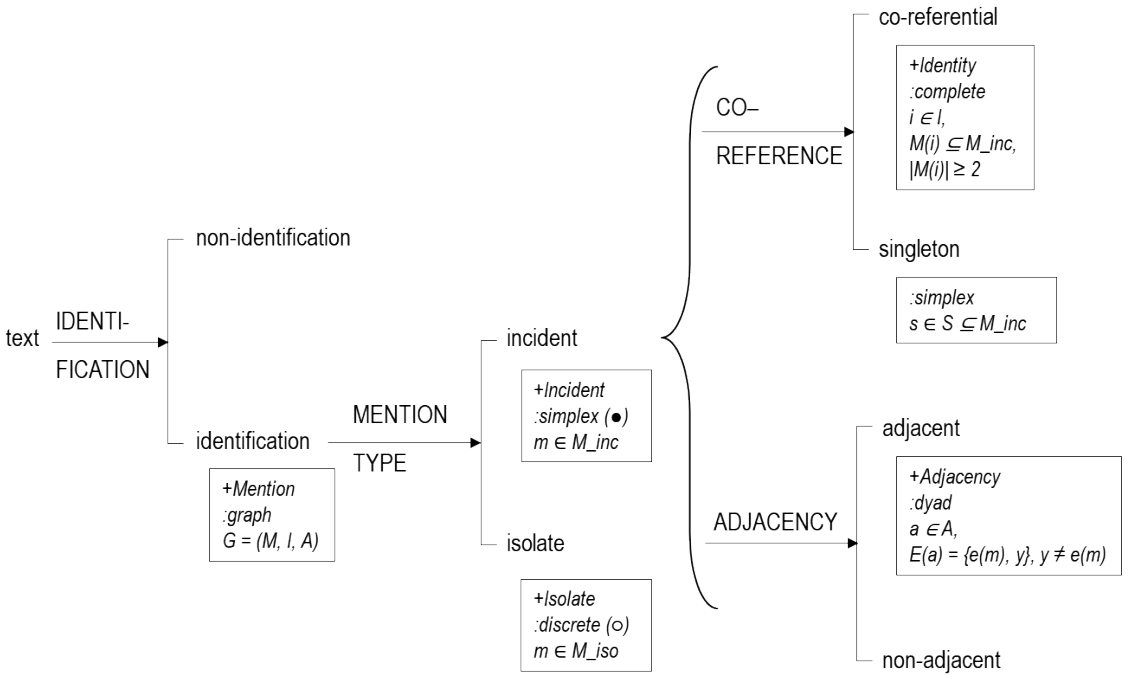
From below, the grammar that goes with a dyad in BP tends to be a clause, or a nominal group with a Qualifier or a personal Deictic.

Table 12: Grammar of mentions in Text 6

line	wording	grammar	feature
1	*a aldeia de Bourg la-Reine* 'the village of Bourg la-Reine'	nominal group: *a aldeia de Bourg la-Reine onde seu pai era prefeito* 'the village of Bourg la-Reine where his father was mayor' Thing *aldeia*, [clausal] Qualifier *onde seu pai era prefeito*	[adjacent]
2	*seu pai* 'his father'	nominal group: personal Deictic *seu*, Thing *pai*	[adjacent]
9	*Cauchy*	clause: *entregando-o a Cauchy* 'handing it to Cauchy'	[adjacent]
13	*o pai* 'the father'	nominal group: specific Deictic *o*, Thing *pai*, [clausal] Qualifier *que, devido a intrigas clericais, se suicidou* 'who, owing to clerical intrigues, killed himself	[adjacent]
21	*o artigo* 'the paper'	clause, no other nominal group: *foi perdido* 'was lost'	[non-adjacent]
22	*as causas* 'the causes'	nominal group: *as causas da Revolução de 1830* 'the causes of the Revolution of 1830': Thing *causas*, [phrasal] Qualifier *da Revolução de 1830*	[adjacent]

The feature [adjacent] inserts the function Adjacency (+Adjacency), and [non-adjacent] inserts nothing. An Adjacency is a function – the element of discourse semantic structure that meets at least two identities, or identities and a singleton, given at once. The identity GALOIS and the identity CAUCHY, for instance, are joined by one adjacency in *entregando-o a Cauchy* 'handing it to Cauchy' (9).

Figure 14: The system of IDENTIFICATION in Brazilian Portuguese.



Each feature preselects a class of graph as follows:

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Table 13: Features, functions and classes of graph in ADJACENCY

feature	**realisation**	**in the graph**	**notation**
[adjacent]	+Adjacency: dyad	edge meeting non-equivalent elements, each identity or singleton ●; class dyad	$a \in A$, $E(a) = \{e(m), y\}$, $y \neq e(m)$
[non-adjacent]	nothing	full node ●, no adjacency; class simplex, inherited from +Incident	none

In notation, 'A' is the set of adjacencies and $S \subseteq M_{inc}$ the singletons. 'E(m)' is the element of a mention 'm'. It is the identity 'i' with $m \in M(i)$, or 'm' itself if 'm' is a singleton. A mention that selects [adjacent] inserts an adjacency $a \in A$ that meets $E(a) = \{e(m), y\}$, where 'y' is another identity $i \in I$ or a singleton $s \in S$, and $y \neq e(m)$. Adjacency preselects a dyad – non-equivalent clusters taken as one element.

CO-REFERENCE and ADJACENCY are simultaneous. A mention that selects [incident] makes a free choice in each, and the two choices combine in four ways (Table 14).

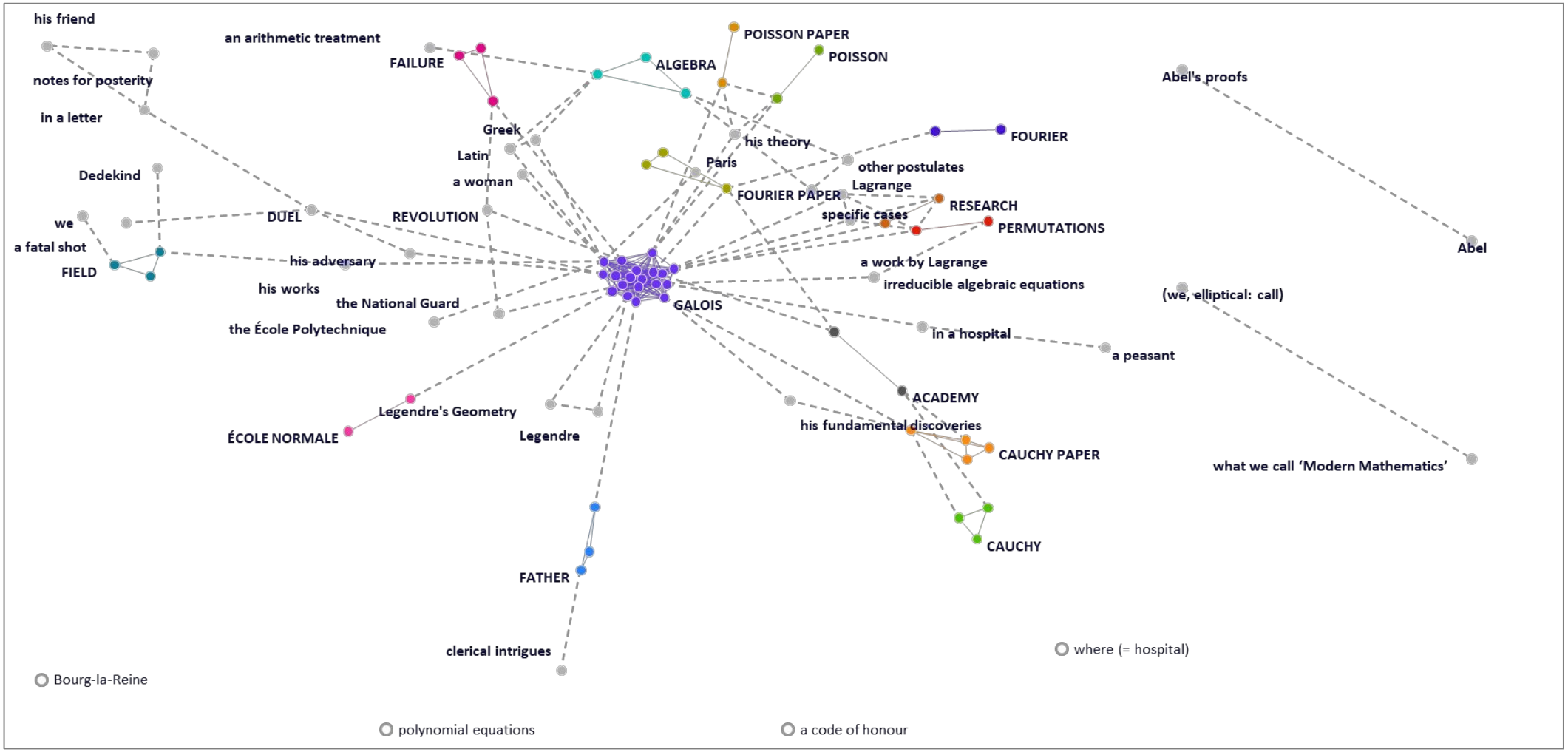
Table 14: Co-selection of CO-REFERENCE and ADJACENCY in Text 6

features	**realisation**	**in the graph**	**example**
[co-referential, adjacent]	+Identity, +Adjacency	node in identity; dyad between that identity and another element	*Cauchy* (9)
[singleton, adjacent]	+Adjacency	lone full node ●; dyad between it and another element	*a Escola Politécnica* (5)
[co-referential, non-adjacent]	+Identity	node in identity; no dyad	*o artigo* (21)
[singleton, non-adjacent]	nothing	lone full node ●; no dyad	*uma roupa* (Table 2, b)

Figure 15 shows that one identity-cluster, GALOIS, is far larger than any other. Around it are mostly singletons and identities of a few mentions, such as *Latim* 'Latin' (3), *a Escola Politécnica* 'the École Polytechnique' (5) and *a guarda nacional* 'the national guard' (24), each joined to it by an adjacency of its own, and CAUCHY, FOURIER and POISSON, each joined to it and to the paper.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Figure 15: The graph of Text 6.



SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Comparably, Text 6 shows more adjacency than co-reference, since many episodes in Galois's life involved different people and events. Due to forces of attraction and repulsion, the graph takes the shape of a spider web.

This graph and that of the folk tale (Figure 13) differ in shape because the two genres differ in what their stages ask of IDENTIFICATION. A narrative resolves a complication through what its characters do to one another, so its two main characters are large identities, joined to each other and to the things between them. A biographical recount records a life, so one identity, its subject, has mentions in most of the text, each other element joined to it by an adjacency of its own and few to one another.

We can put forward a testable hypothesis, that the shape of the graph is a feature associated with the genre. Genres organised around interaction among a few central identities, such as plotted narrative, tend towards several large identities joined to one another, and adjacencies between several clusters. Genres organised around one subject, such as biography, tend towards one large identity with the other elements around it, mostly adjacent to it alone.

5 Homophora and exophora

This section begins with Text 7, which is another text of the biographical recount genre. It also records the life of Galois.

TEXT 7 [PB_BIO_02]

Orientation	1	Galois é natural de Bourg-la-Reine, perto de Paris.	Galois is a native of Bourg-la-Reine, near Paris.
Record of events	2	Na escola o jovem tinha um aproveitamento muito irregular,	At school the young man had a very irregular record,
	3	embora conseguisse ler com facilidade matemáticos importantes como Joseph-Louis Lagrange, Adrien-Marie Legendre, Augustin-Louis Cauchy e Friedrich Gauss.	although he managed to read with ease important mathematicians such as Joseph-Louis Lagrange, Adrien-Marie Legendre, Augustin-Louis Cauchy and Friedrich Gauss.
	4	Há registros do professor de retórica reclamando	There are records of the rhetoric teacher complaining
	5	que, aos 16 anos, era inútil fazê-lo se interessar por qualquer disciplina:	that, at 16, it was useless to make him take an interest in any subject:
	6	"Dominado por sua paixão pela matemática, ele descuidou inteiramente de tudo o mais".	"Dominated by his passion for mathematics, he entirely neglected everything else."
	7	Foi essa paixão que o levou a uma grande ambição:	It was this passion that led him to a great ambition:
	8	achar um modo de resolver equações de grau cinco.	to find a way of solving equations of degree five.
	9	Ele também desejava entrar para a Escola Politécnica, a principal instituição de ensino superior do país,	He also wished to enter the École Polytechnique, the country's leading institution of higher education,
	10	e tentou duas vezes sem conseguir.	and tried twice without succeeding.
	11	Segundo os estudiosos de sua vida,	According to specialists in Galois's life,
	12	é provável que ele não estivesse preparado para isso.	it is likely that he was not prepared for it.
	13	Uma das reclamações de seus examinadores é a de	One of his examiners' complaints is
	14	que ele fazia boa parte dos cálculos de cabeça	that he did a good part of the calculations in his head

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

	15	e dava apenas os resultados,	and gave only the results,
	16	sem demonstrar a resolução do problema.	without showing the solution of the problem.
	17	Isso deixava incrédulo – e desagradava – quem o inquiria.	This left whoever questioned him incredulous – and displeased.
	18	Optou, então, pela Escola Preparatória, nome temporário dado à Escola Normal.	He then opted for the Preparatory School, a temporary name given to the École Normale.
	19	A militância de Galois a favor da República numa França monarquista o levou à expulsão da instituição e à sua prisão por duas vezes.	Galois's militancy in favour of the Republic in a monarchist France led him to expulsion from the institution and to his imprisonment on two occasions.
	20	Apaixonou-se por Stéphanie Potterin du Motel	He fell in love with Stéphanie Potterin du Motel
	21	e foi morto em um duelo não se sabe exatamente por quem.	and was killed in a duel, it is not known exactly by whom.
	22	Uma das versões diz	One of the versions says
	23	que o desafiante teria sido alguém próximo da moça.	that the challenger was someone close to the young woman.
	24	Outra afirma	Another claims
	25	que foi uma maquinação de monarquistas.	that it was a scheme of monarchists.
	26	E há ainda uma terceira contando	And there is also a third, telling
	27	que o próprio Galois teria provocado sua morte	that Galois himself had brought about his death
	28	para insuflar uma rebelião contra o rei Carlos X.	to stir up a rebellion against King Charles X.
	29	A única certeza é	The only certainty is
	30	que ele foi atingido por um tiro na barriga	that he was hit by a shot in the belly
	31	e morreu no dia 30 de maio de 1832, depois de 12 horas.	and died on 30 May 1832, 12 hours later.
	32	Galois escreveu cinco trabalhos pequenos e três memórias.	Galois wrote five small works and three memoirs.
	33	No total, sua obra matemática tem 60 páginas.	In total, his mathematical work has 60 pages.
	34	Em vida, apenas os artigos curtos foram publicados.	In his lifetime, only the short papers were published.
	35	Depois de sua morte, a mãe repassou para um amigo do jovem, Auguste Chevalier, vários manuscritos e três cartas.	After his death, his mother passed on to a friend of the young man, Auguste Chevalier, several manuscripts and three letters.
	36	Duas eram sobre política,	Two were about politics,
	37	mas uma continha o resumo das memórias [[que ele havia escrito]]	but one contained the summary of the memoirs [[that he had written]]
	38	e foi publicada na revista Revue Encyclopédique, em setembro de 1832.	and was published in the journal Revue Encyclopédique, in September 1832.
	39	Apenas em 1846 todos os trabalhos foram publicados no Journal de Mathématiques Pures et Appliquées.	Only in 1846 were all the works published in the Journal de Mathématiques Pures et Appliquées.

'From above', the generic structure of Text 7 is Orientation ^ Record of events. In Orientation, Galois, Bourg-la-Reine and Paris are joined to one another in (1). In the Record of events, Galois is mentioned along with others that do not form an identity with Galois: *Stéphanie Potterin du Motel* (20), *o rei Carlos X* 'King Charles X' in (28) and *Auguste Chevalier* in (35). Others are mentioned more than once, for example *A Escola Preparatória* 'the Preparatory School' and *a Escola Normal* 'the École Normale' in (18).

From roundabout, we apply commutation on a dyad as we did for Text 6 (examples in Table 15). In the first two rows it changes the dyad and in the last none. *Stéphanie Potterin du Motel* (20) and *o rei Carlos X* (28) are [adjacent]; *A única certeza* (29) is [non-adjacent].

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Table 15: Commutation in Text 7

line	original	commuted	change elsewhere
20	*Stéphanie Potterin du Motel*	*Lagrange*	Galois identity-cluster joined to *Lagrange*, not *Stéphanie Potterin du Motel*
28	*o rei Carlos X* 'King Charles X'	*a República* 'the Republic'	*uma rebelião* 'a rebellion' joined to *a República*, not *o rei Carlos X*
29	*a única certeza* 'the only certainty'	*Poisson*	none: element joined to no other

From below, the grammar that goes with a dyad in Text 7 (Table 16) patterns with that of Text 6.

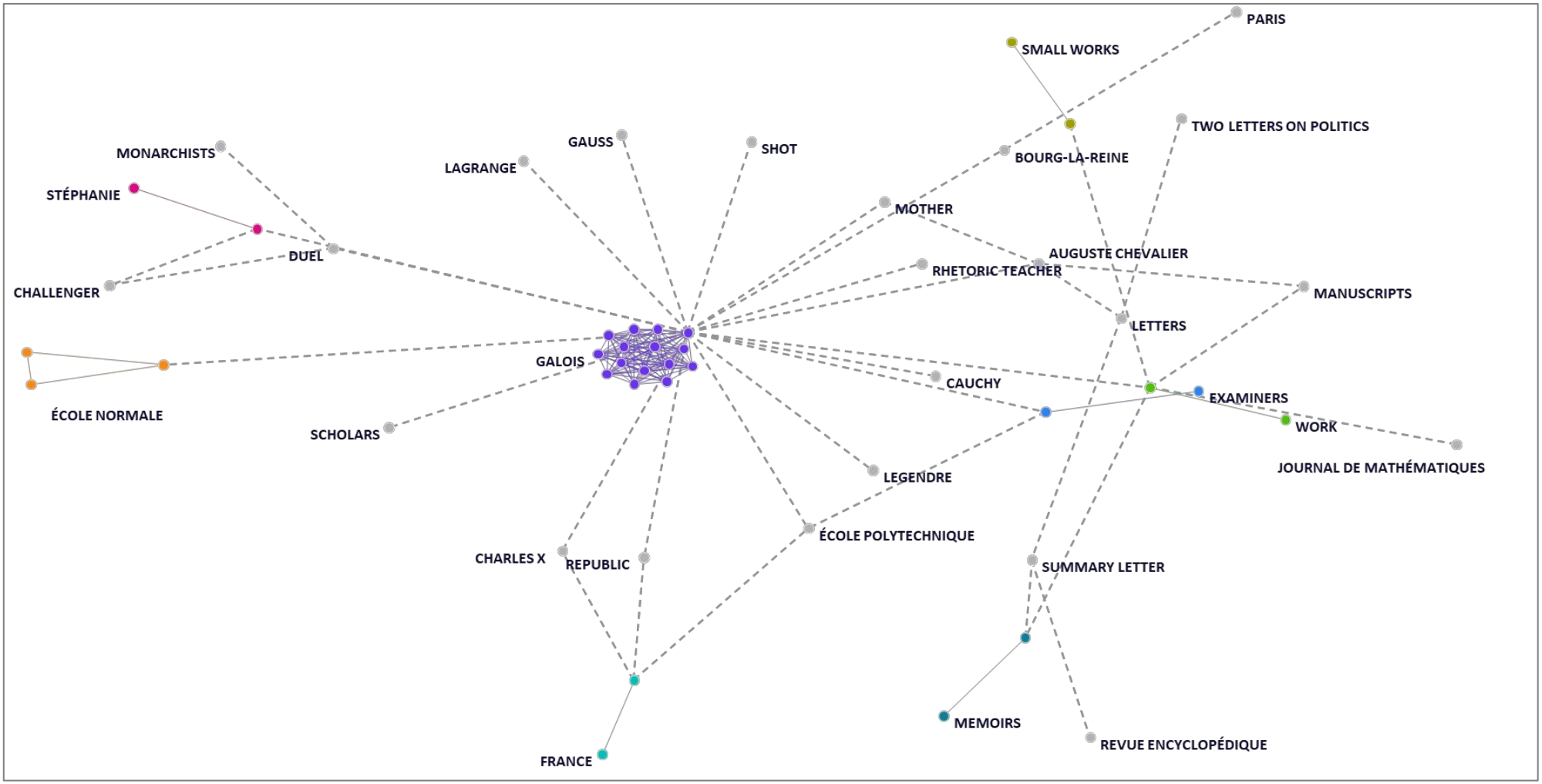
Table 16: Grammar of mentions in Text 7

line	wording	grammar	feature
1	*Bourg-la-Reine*	clause: *Galois é natural de Bourg-la-Reine, perto de Paris* 'Galois is a native of Bourg-la-Reine, near Paris'; Galois, Bourg-la-Reine, Paris single nominal groups	[adjacent]
11	*os estudiosos de sua vida* 'the scholars of his life'	nominal group: Thing *estudiosos*, [phrasal] Qualifier *de sua vida*, personal Deictic *sua* 'his'	[adjacent]
22	*uma das versões* 'one of the versions'	single nominal group: *Uma das versões diz* 'One of the versions says'	[non-adjacent]

In Figure 16, one identity-cluster, GALOIS, has 36 mentions. The next largest, the École Normale, meets three mentions (19); fifteen other identities meet two. GALOIS is joined by an adjacency of its own to 36 elements, among them *Bourg-la-Reine* and *Paris* (1), *a Escola Politécnica* (9), *Stéphanie Potterin du Motel* (20), *um duelo* 'a duel' (21) and *um tiro* 'a shot' (30). The whole text is one large identity at the middle, with elements joined to it and small groups joined to one another around it.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Figure 16: The graph of Text 7.



SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

The graph of Text 7 on its own leaves something open, though. *Bourg-la-Reine* and *Paris* (1), and *Lagrange*, *Legendre*, *Cauchy* and *Gauss* (3), are each a singleton in the graph of Text 7. A singleton says that no other mention in its own text is the same as it. Martin (1992) explains the places from which the identity of a phoric item, such as these singletons, can be recovered: the non-verbal context of situation (exophora) and the context of culture (homophora). Proper names are among the examples of homophora.

The graph of a text says which mentions are the same as which, on the evidence of commutation, which does not refer to where the mentions are found. Text 6, the other biographical recount of Galois, has *a Escola Politécnica* 'the École Polytechnique' (5) as a singleton joined by an adjacency to the identity GALOIS; Text 7 also mentions the École Polytechnique (9). Interestingly, commutation applies to the two texts together (Table 17).

Table 17: Commutation between mentions from Texts 6 and 7

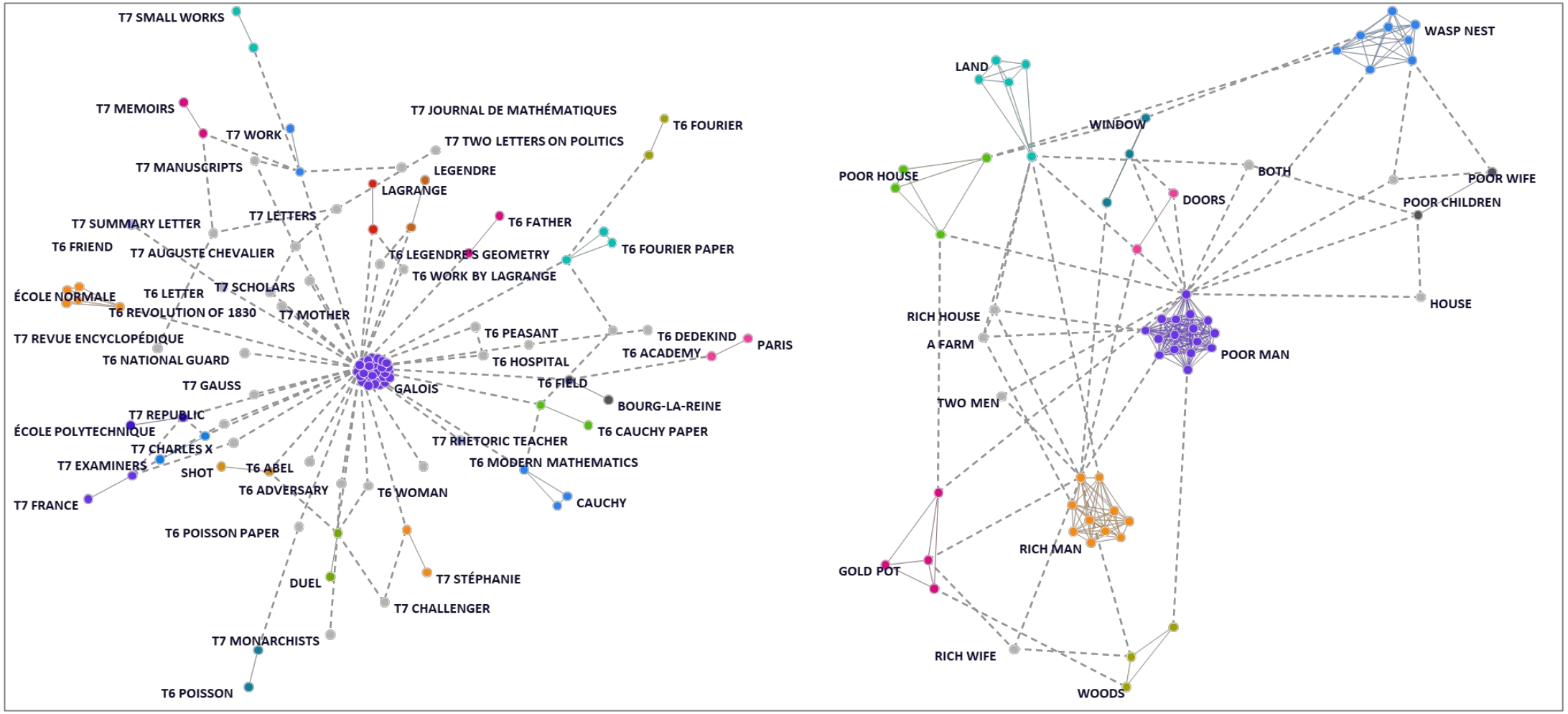
lines in Text 6	**source**	**commuted**	**change in Text 7**
Text 6, 1	*Évariste Galois*	*Cauchy*	*Galois* (Text 7, 1) other Galois mentions in Text 7 not same as Text 6 (1)
Text 6, 5	*a Escola Politécnica*	*a Escola Normal* (Text 6, 14)	*a Escola Politécnica a principal instituição de ensino superior do país* (Text 7, 9) not same as Text 6 (5)
Text 6, 5	*a Escola Politécnica*	*a principal instituição de ensino superior do país* (Text 7, 9)	none: *a Escola Politécnica* (Text 7, 9) still same

Commutation on Text 6 changes mentions in Text 7 in the first two rows and none in the last: the contrast of CO-REFERENCE, across two texts. One identity meets *a Escola Politécnica* (Text 6, 5) and the two mentions in Text 7 (9), as it would within one text. *Uma mulher* 'a woman' (Text 6, 35) and *Stéphanie Potterin du Motel* (Text 7, 20) are not met by one identity: neither text states that they are the same.

Figure 17 draws G = (M, I, A) over the mentions of Texts 6 and 7 together. For the sake of comparison, we also added Text 5, which shares no mention with either, to observe the shape given by the force layout.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Figure 17: The graph of a corpus: Texts 5, 6 and 7.



SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

In Figure 17, Texts 6 and 7 form one cluster around GALOIS. Each identity with mentions in both texts is one element of the graph, and a dyad inserted in both texts is one adjacency: Galois and Bourg-la-Reine, Galois and Paris, and Bourg-la-Reine and Paris (1 in both), Galois and the École Polytechnique (Text 6, 5; Text 7, 9), and Galois and the duel (Text 6, 35; Text 7, 21). Other dyads occur in one text only: Galois with Cauchy, Fourier and Poisson in Text 6, and Galois with the Republic, the militancy and the imprisonment in Text 7.

Bourg-la-Reine, Paris, Lagrange, Legendre and Cauchy are singletons in Text 7. But they are met by an identity with mentions in Text 6 – thus, they are no longer singletons. What we call homophora is actually singletons with an identity in another text. The same happens to *a Escola Politécnica* (Text 6, 5). Consequently, Martin's recovery of an identity from the context of culture, homophora, needs no relation of its own. It is actually co-reference across mentions in different texts, tested by the same commutation, with no new edge type, closure property or realisation statement. Identity is thus tested against the mentions of the texts, not assumed from 'shared knowledge'.

By contrast, Text 5 stands as a separate cluster to the side, pushed apart by the force layout, with no identity or adjacency between it and Texts 6 and 7.

The homophora result extends as a hypothesis to Martin's exophora, where the identity is recovered from the non-verbal context of situation (*Pass me that knife, will you?*): the same relation with a non-verbal text. Commutation does not refer to the modality of the text, so a mention in an image, or an object in the scene, is replaced as any mention is.

6 Conclusion

Identification in Brazilian Portuguese can be described as three systems, MENTION TYPE, CO-REFERENCE and ADJACENCY, realised by one structure, the identity graph G = (M, I, A). The graph is given at once, so it is a structure of indefinite extent and has no rank. It is the covariate structure for discourse semantics first identified by Lemke and Martin. In addition, our modelling carries through Gleason's (1968) finding that participant identification belongs to the whole text and Martin's (1992) synoptic realisation statements, and makes them annotatable by hand and comparable across texts and, in principle, across languages.

The results enabled us to answer the four driving questions posed in Section 1. (1) The entry condition of the system is the text. (2) The structure it generates is a graph, given at once, so the 'travelling' nature of text needs no place in it. (3) Society and culture, the mind, or the "outside the text" are left aside: the description looks only at genre from above, not at people in society. (4) And because mentions, identities and adjacencies are stated explicitly, annotation can be automated, producing the tables and graphs of this paper.

More specifically, the paper could establish that co-reference is an equivalence relation, decided by commutation, so each mention has one identity independently of any particular text. Retrieval relations defined by serial order then belong to the description of an instance and need no mechanism in the system. In turn, homophora becomes co-reference across mentions in different texts. Following this principle, a singleton in its own text can be met by an identity with mentions in another, as *a Escola Politécnica* 'the École Polytechnique' and Paris are in Texts 6 and 7. A text is limited by its genre; the identity is not, and its boundary

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

belongs to the text. This suggests that the indefinite extent of discourse is not bound or constrained by the limits of a text. Discourse sprawls over, connecting meanings in different texts. Likewise, exophora is hypothesised to be the same relation with a non-verbal text (which presently remains untested).

The graphs of the texts analysed differ in shape: discrete for the Description of Text 4, several large identities joined to one another for Text 5, one large identity, Galois, with the other elements around it for Texts 6 and 7. Graph shape is associated with genre, which relates a discourse semantic pattern to genre from above, without appeal to the unfolding of any instance.

Our contribution to typological work rests on the fact that the description is systemic and axial, so it can be carried to other languages. A great deal of identification description has been done privileging its realisation in grammar. The typological implication is the workload of identification divides differently between clause and nominal group across languages, as Martin (1983) shows for English, Tagalog and Kâte and Martin and Cruz (2025) for Filipino, where it is bound up with the verb's focus and voice system. But, again, these are differences in grammatical realisation.

If the view "from roundabout" is priviledged, as it is in this paper, then stated system with a replicable annotation procedure makes that variation comparable – and not only in terms of how identification is realised in the grammar of different languages, but also relative to discourse structures in their own right and "from above" in the relation between those structures and genres of different languages. Finally, the graph is explicit enough to be annotated by hand and checked by commutation – thus all data can be audited in every step of the way. We are not assuming any category (system, structure, class, or graph) that can not be annotated by anyone – including machines – and obtained following the same methodological steps.

Finally, it is fundamental to state that the paper left several matters open, and could be picked up by future research.

(i) Adjacency is read off the text by commutation; no rule here generates it. From above it is accounted for by genre – the zoo recount takes the writer past a hippopotamus – but which stages bring which joins is unknown.

(ii) A set and its members, as in *Havia dois homens, um rico e outro pobre* 'There were two men, one rich and one poor' (Text 5, 1), stand in a relation the three systems do not state: *dois homens* is not the same as *um rico*, so it is not co-reference, yet *ambos* 'both' (8) is the same as *dois homens*. Although this has no implication to the description of BP, it may be relevant to other languages.

(iii) Vocatives, as *Ó compadre* 'Friend' (Text 5, 25), pass the same test as third-person mentions, so the narrator's and the characters' mentions sit in one graph undistinguished. We may wonder whether interpersonal meanings, such as a Vocative, can form co-reference or adjacent relations.

(iv) The description awaits comparison with alternative tracking models – which also view discourse as different from an overgrown clause; with no rank, composition, co-variate and of indenfinite extent.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

(v) The testing of the association between graph shape and genre in other corpora and languages is needed.

(vi) How patterns of mentions associate across texts belongs to instantiation, as probabilities, or to individuation, as bond complexes, depending on which texts are compiled in a given corpus, or are available to a given reader to individuation. These issues have not been addressed here. We observed them empirically, but we did not formulate them theoretically. Some theoretical piece explaining how meaning in discourse accretes over time, dynamically, which then can be associated with how bonds sprawl through texts to build communities is urgently needed.

Ethical and Legal Declarations

Corresponding author
Giacomo Figueredo (giacomo.figueredo@ufop.edu.br)

Funding
This project was financed by the National Council for Scientific and Technological Development (CNPq/Brazil), process no. 310547/2023-7.

Conflict of interest / Competing interests
The author declares no conflict of interest.

Research involving human participants and/or animals
This study did not involve human participants, human data, or animal subjects requiring ethical approval; it is a corpus-based linguistic description drawing on previously compiled, free-access corpora.

Informed consent
Not applicable.

Declaração de disponibilidade de dados de pesquisa / Research data availability statement
Data – including annotation, corpus, calculations, and graphs – will be made available on request.

Author contributions
Giacomo Figueredo is the sole author and is responsible for all aspects of the manuscript (conceptualisation, data curation, formal analysis, methodology, writing).

Declaração de uso de Inteligência Artificial / Use of generative AI / Large Language Models
A generative AI tool (Claude, Anthropic) was used as a software tool for double-checking language proofreading, formatting, reference checking, and generating the graph visualisations in Section 3.3 from the layout parameters specified in the paper and the data in the manually annotated tables.

Acknowledgements
We thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for funding this research.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

References

Bondy, J. A. & U. S. R. Murty. 2008. *Graph theory* (Graduate Texts in Mathematics 244). New York: Springer.

Catford, J. C. 1965. *A linguistic theory of translation: An essay in applied linguistics*. London: Oxford University Press.

Doran, Y. J. in press. Modelling context from multiple perspectives. *Language, Context and Text*.

Figueredo, G. 2014. Uma metodologia de perfilação gramatical sistêmica baseada em corpus. *Letras & Letras* 30(2). 17–45.

Figueredo, G. 2020. Axial argumentation and cryptogrammar in textual grammar. In J. R. Martin, Y. J. Doran & G. Figueredo (eds.), *Systemic functional language description: Making meaning matter*, 129–161. London/New York: Routledge.

Figueredo, G. 2021. The nominal group in Brazilian Portuguese. *WORD* 67(4). 461–492. https://doi.org/10.1080/00437956.2021.1993588

Figueredo, G. & G. Freitas. 2026. The verbal group system and structure in Brazilian Portuguese. *Lingua* 329. Article 104070. https://doi.org/10.1016/j.lingua.2025.104070

Fontaine, L., K. Jones & D. Schönthal. 2023. *Referring in language: An integrated approach*. Cambridge: Cambridge University Press.

Fruchterman, T. M. J. & E. M. Reingold. 1991. Graph drawing by force-directed placement. *Software: Practice and Experience* 21(11). 1129–1164.

Gleason, H. A., Jr. 1965. *Linguistics and English grammar*. New York: Holt, Rinehart & Winston.

Gleason, H. A., Jr. 1968. Contrastive analysis in discourse structure. *Monograph Series on Languages and Linguistics*.

Hagberg, A. A., D. A. Schult & P. J. Swart. 2008. Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught & J. Millman (eds.), *Proceedings of the 7th Python in Science Conference (SciPy2008)*, 11–16.

Halliday, M. A. K. 2005. Fuzzy grammatics: A systemic functional approach to fuzziness in natural language. In M. A. K. Halliday & J. J. Webster (eds.), *Computation and quantitative studies* (Collected Works of M. A. K. Halliday, vol. 6), 213–238. London: Continuum.

Halliday, M. A. K. & W. S. Greaves. 2008. *Intonation in the grammar of English*. London: Equinox.

Hao, J. 2020. *Analyzing scientific discourse from a systemic functional linguistic perspective: A framework for exploring knowledge building in biology*.

Hao, J. 2026. Rethinking protolanguage. Plenary paper, ASFLA 2026 Conference. Australian Systemic Functional Linguistics Association.

Harary, F. 1969. *Graph theory*. Reading, MA: Addison-Wesley.

Hjelmslev, L. 1961. *Prolegomena to a theory of language*, 2nd edn. Transl. Francis J. Whitfield. Madison: University of Wisconsin Press.

Lemke, J. L. 1985. Ideology, intertextuality, and the notion of register. In J. D. Benson & W. S. Greaves (eds.), *Systemic perspectives on discourse*, vol. 1.

Martin, J. R. 1983. Participant identification in English, Tagalog and Kâte. *Australian Journal of Linguistics*.

Martin, J. R. 1985. Process and text: Two aspects of human semiosis. In J. D. Benson & W. S. Greaves (eds.), *Systemic perspectives on discourse*.

Martin, J. R. 1992. *English text: System and structure*. Amsterdam: John Benjamins.

SciELO Preprints - This document is a preprint and its current status is available at: https://doi.org/10.1590/SciELOPreprints.18897

Martin, J. R. 2008. Tenderness: Realisation and instantiation in a Botswanan Town. *Odense Working Papers in Language and Communication* 29. 30–58.
Martin, J. R. 2015. Meaning beyond the clause: Co-textual relations. *Linguistics and the Human Sciences* 1(2-3). 203–235. https://doi.org/10.1558/lhs.34711
Martin, J. R. 2026. Secular communion: So much trouble in the world. *Estudios de Lingüística Aplicada* 44(83). 15–43. https://doi.org/10.22201/enallt.01852647p.2026.83.1146
Martin, J. R. & P. A. T. Cruz. 2025. Textual meaning: Organising information flow. In *Analyzing Filipino discourse*.
Martin, J. R. & S. J. Dreyfus. 2015. Scaffolding semogenesis: Designing teacher/student interactions for face-to-face and online learning. In S. Starc, C. Jones & A. Maiorani (eds.), *Meaning making in text: Multimodal and multilingual functional perspective*, 265–298. London: Palgrave Macmillan.
Martin, J. R. & D. Rose. 2007. *Working with discourse: Meaning beyond the clause*.
Martin, J. R. & P. R. R. White. 2005. *The language of evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan.
Novák, M., M. Konopík, A. Nedoluzhko, M. Popel, O. Pražák, J. Sido, M. Straka, Z. Žabokrtský & D. Zeman. 2026. Findings of the fifth shared task on multilingual coreference resolution: Expanding datasets for long-range entities. In *Proceedings of the 5th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI-CRAC 2026)*. https://aclanthology.org/2026.codi-1.21
Quiroz, B. 2026. From lexicogrammar to context: A systemic functional perspective on language description illustrated through reference in Spanish. Plenary paper, International Systemic Functional Congress, Hong Kong.
Rose, D. 2019. *Reading to learn: Accelerating learning and closing the gap*. Sydney: R2L.
Whorf, B. L. 1945. Grammatical categories. *Language* 21(1). 1–11.

This preprint was submitted under the following conditions:

- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a Creative Commons CC-BY license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.