# Category theory, linguistics and pronouns

निरंजन*
Department of Linguistics
University of Mumbai

PhD coursework examination
Paper I: Linguistic analysis
Term-paper submission 2
January 30, 2024

**Abstract**

This article first discusses the essentials of determining a linguistic category and then we examine whether pronouns can be called a category or not. I propose that pronouns do stand as a separate linguistic category because of the features they show.

## Contents



## 1 Introduction

In mathematics, we abstract things, to be able to make generalised claims about them. Abstracting involves focusing on some aspects of a phenomenon while ignoring some. When some details are ignored, it becomes easier to apply logical operations on them. Like in mathematics, abstraction is seen in linguistics too.

The production of language ('performance' in Chomskyan terms) is comprised of a lot of variation which is generally ignored by linguists while profiling a language, e.g., while describing the phonology of English, [t] sound is recognised as a phoneme, but the nature of it in the word 'tooth' is very different from it in the word 'table'. Both of them are still said to be the instances of the /t/ phoneme. This is an example of abstraction. The ignored differences aren't unnecessary, they are just irrelevant for the description of the features which we are trying to focus on.

Let's take another example from phonology. The bunch of sounds that are articulated in the frontal area of the mouth are called 'anterior'. They consist of sounds with varying places and manners of articulation, but we may have a stricter grouping by adding more conditions. E.g., instead of just anterior, we may have a group of anterior plosives where the number of sounds will be lesser. Similarly we can reduce it to anterior voiced plosives. To understand this visually, let's have a look at figure 1[I].

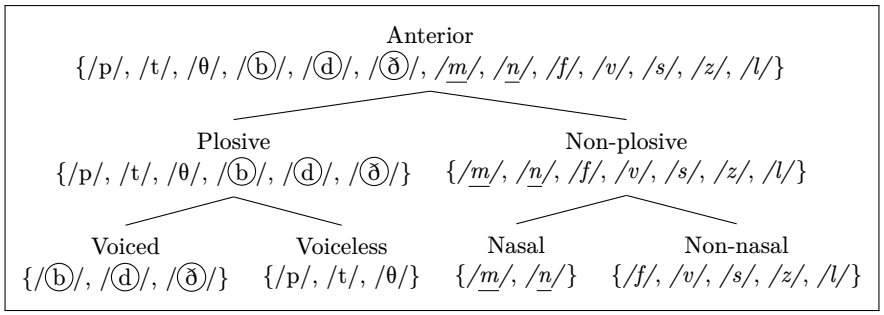


Figure 1: Categorisation of anterior sounds of English

As can be seen in the figure, when we add conditions, the number of sounds reduce. But this isn't just about the membership of sounds under certain groups. An essential of categories is the relationship between the objects, i.e., morphisms. Even in this example, one could easily identify that there are nasals which actually are voiced, but the tree structure cannot express the voiced feature of them. In figure 1 we have a set-theoretic visualisation of the anteriors, let's have a look at figure 2 for a more nuanced depiction of the relations between these groups.

In figure 2, we can see the following sounds grouped at the respective nodes. Now interestingly, (c), (e), (g) and (h) nodes can host non-anterior sounds too.

**(a):** /b/, /d/, /ð/

**(b):** /m/, /n/, /v/, /z/, /l/

**(c):** /ʤ/, /g/

[I]Upright letters represent voiceless anterior plosives, circled ones represent voiced anterior plosives, underlines represent anterior nasals and lastly oblique letters represent non-plosive non-nasal anteriors.

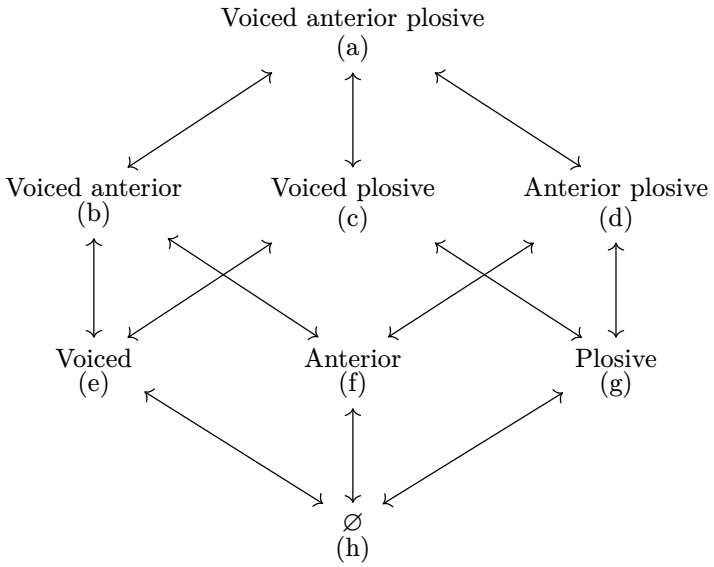


Figure 2: Commutative diagram of anterior sounds of English

**(d):** /p/, /t/, /θ/

**(e):** /ʒ/, /ŋ/, /r/, /j/, /w/

**(f):** /f/, /s/

**(g):** /ʧ/, /k/

**(h):** /ʃ/, /h/

The (b) node now hosts nasals and voiced fricatives. It is directly connected to (e), (f) and (a) which respectively host voiced non-anterior non-plosives, anterior voiceless non-plosives and voiced anterior plosives. In figure 1, there was no way to see the connection between the node hosting nasals and the node hosting voiced anterior plosives. But the morphisms (depicted by the arrows in the diagram) allow us to capture these relations. Note that similar commutative diagrams can be drawn for separate nodes too. E.g., the (b) and (e) nodes can have further categorisation such as sibilants, approximants. All these objects share very specific properties which are shared seen in the objects to which they are connected with morphisms. This is an important difference between categories and sets. A set of two unrelated elements is possible, but a category of two unrelated objects is not. Categorisation helps us make predictions regarding the structure of the phenomenon, sets aren't much equipped to do this.

# 2 Pronouns

The discussion so far aims to demonstrates how linguistic phenomena can undergo mathematical categorisation. In order to verify whether pronouns constitute a category, we will first assume that it is not a category. We will first explore the features of forms which are typically called pronouns. We will mathematise them in order to see the morphisms and then conclude.

## 2.1 Nouns vs. pronouns

Often, pronouns are believed to be contained in the category of nouns. Since nouns and pronouns share the core grammatical features, such as person number and gender, but there are interesting patterns which demand some attention.

Let's have a look at the following Marathi data. Marathi is a gender-rich language. It has three genders and its verbs agree with the gender of the non-cased phrases.

(1) a. माकड म्हणालं मी आंबे खातो.

makəɽ mʱəɳ̆-a-l-ə mi amb-e kʰa-ʈ̪-o
monkey.NSG say-AUG-PFV-N 1SG mango-PL eat-IPFV-MSG

A monkey said, 'I eat apples'.

b. बाळा, तू किती छान दिसतेस!

baɭ̆-a ʈ̪u kiʈ̪i ʧʰaɳ̪ ɖis-ʈ̪-e-s
kid.NSG-VOC 2SG how good look-IPFV-FSG-2

Kiddo, how pretty you look!

As can be seen in 1a, 'माकड' (/makəɽ/) is a non-cased neuter noun which triggers neuter agreement on the verb, i.e., 'म्हणालं' (/mʱəɳ̆-a-l-ə/). Similarly in 1b, 'बाळ' (/baɭ/) is a neuter noun which is seen in its vocative form. What is interesting about these examples is, neither of them show neuter agreement on verbs in the embedded clauses. Marathi doesn't seem to allow neuter agreement in the first and second person.

What does it mean from the perspective of categories? Since all nominals of Marathi show gender, its an essential categorical feature. In Marathi, the gender feature has three values, but in the case of first and second person, it has only two values. This becomes a crucial difference. In terms of categories, the objects contained in the category of nouns always undergo a three-way split of gender-values. But the nouns contained in the category of first and second person pronouns undergo only a two-way split. Note that the same pattern is found in other few Indo-Aryan languages having 2+ genders (i.e., Konkani and Gujarati[2]).

[2]Kannada and Sanskrit too have 2+ genders, but their verbs don't agree for the gender feature, hence they are excluded.

## 2.2 Dynamic reference

It is a well known fact about pronouns that their reference is computed dynamically. E.g., the person referred by a pronoun is always understood in context. Table 1 describes the referential nature of all the three pronouns. Users put themselves in an analytic picture, called the *'origo'* (Abraham, 2012).

| | |
|---|---|
| First person | The speaker |
| Second person | The addressee |
| Third person | Not the speaker, nor the addressee |

Table 1: Dynamic references of pronouns

Note that the third person has the most loose conditions. It is so because everything except first and second person pronoun has to be in the third person. We know that morphologically too. We discussed how semantically persons are mapped so that we can discuss the dynamic assignment of references.

As Abraham (2012, p. 216, figure 2) shows, the references exist only in the perceptions of people involved in the communication. Not just the speaker, but the addressee too computes the reference dynamically. Suppose $x$ and $y$ are conversing, and $x$ is speaking, the dynamic assignment of reference is not only in $x$'s head, but in $y$'s head too. Suppose $x$ says, 'I will give him a pen.', $y$ too understands that they aren't referred to by $x$. Suppose, then, $y$ responds, 'I gifted him a book yesterday.', it's not just a continuation of the earlier *origo*. Now they have switched the roles and put themselves as the speaker. Thus it should be noticed that this dynamicity is discontinuous.

These features, too, become crucial for the categorisation of the pronominal set. Featurally, only first and second person pronouns seem to have the features to denote speech roles. I propose that these two have a PERSPECTIVE feature. This feature causes them to be discontinuously dynamic. It becomes clearer why only these two are proposed to have this feature when we discuss the behaviour of the remainder.

It may feel like third person nouns too are dynamic, since they don't stick to that person value always. There is a possibility that they join the discourse too and become the speaker or the addressee. This is why merely saying a dynamic reference does not suffice. The PERSPECTIVE feature is intuitive here, since that captures this difference pretty well. Third person nouns are the ones which are a part of the discourse since the speaker and addressee are mentioning them, but they are the only nouns which can't have their own *origo*-s. The moment that happens, they shift and take the forms contained in earlier group with PERSPECTIVE feature. This explains why they can have a continuous referential mechanism. Only when nouns are -PERSPECTIVE, they can have continuous dynamic references. In the situation discussed earlier, whether $x$ says *him* or $y$ says so, the person referred to by that form constant. Since it doesn't have its *perspective* (in the context of conversation), the reference keeps on going to the same person. This can't happen to the forms which do have PERSPECTIVE(s).

Abraham (2012) distinguishes between these two as *strong* deixis and *weak* deixis. According to him, the first and second person deixis is strong. He claims that the stronger deixis connects the *origo* to the object directly, whereas the weaker deixis, connects the *origo* to a sign which then connects it to an object (cf. Abraham, 2012, figure 3). This strengthens the proposal of the PERSPECTIVE feature, because in the stronger deixis, the connection is made directly by the members of the *origo*. Whereas in the third person, the members use signs (which featurally are −PERSPECTIVE, but have the ability to point out to other objects). They can only point out things outside the *origo*.

In Marathi, there are some tricky usages of the first person pronoun, which demand a different treatment.

(2) तुझा मी सबळ आहे.

| | | | |
|---|---|---|---|
| t̪udʑʱ-a | mi | səbəɭ | ahe |
| 2.OBL.GEN-AGR | self.M | strong | be.PRS |

Your 'self' is strong.

'मी' (/mi/) is the first person pronoun of Marathi, but there are several reasons to believe that in this case it is a third person noun (which my glossing also attempts to show). Firstly, this can be referred to by third person sequent pronouns. It isn't a part of the *origo*, as the speaker is treating the self as a third person entity which they are describing. Morphologically, first and second person nouns are never gender-conformed. They depend on the gender of the concerned member of the *origo*, but here, as can be seen, it (necessarily) shows masculine gender, even if the addressee is feminine.

# 3 Conclusion

Considering all the data we have seen till now, I infer that the distinction between the first/second and the third person pronouns is very significant. We saw a distinctive feature which enables us to distinguish between the two. These two form separate categories in the language.

## Abbreviations

| | | | |
|---|---|---|---|
| 1 | First person | N | Neuter |
| 2 | Second person | OBL | Oblique |
| AGR | Agreement | PFV | Perfective |
| AUG | Augment | PL | Plural |
| F | Feminine | PRS | Present |
| GEN | Genitive | SG | Singular |
| IPFV | Imperfective | VOC | Vocative |
| M | Masculine | | |

## References

Abraham, W. (2012). Traces of bühler's semiotic legacy in modern linguistics. In W. Abraham & E. Leiss (Eds.), *Modality and theory of mind elements across languages* (pp. 211–250). De Gruyter Mouton. https://doi.org/doi:10.1515/9783110271072.211