Filosofia Unisinos
Unisinos Journal of Philosophy
17(3):320-325, sep/dec 2016
Unisinos – doi: 10.4013/fsu.2016.173.09

PHILOSOPHY SOUTH

Redefining (Katz's) natural languages[1]

Ana Clara Polakof[2]

ABSTRACT

This article proposes a new way of defining Katz's natural languages. Katz defines natural languages as abstract objects, which are later understood as collections of sentences. This is due to the fact that linguistics is understood as a part of mathematics. He proposes a realist linguistics that is not easy to adopt. We think that a realist linguistics that does not consider that linguistics is a part of mathematics could be better embraced. If linguistics were not a part of mathematics, a new definition of natural languages as abstract entities should be provided. We propose to use a hierarchized ontology, and to define natural languages as linguistic states of affairs, which result from the instantiation of a linguistic property in lower level linguistic abstract entities.

Keywords: Katz, Natural Languages, Linguistic Abstract Entities.

Introduction

A realist linguistics is difficult to defend. It is very counterintuitive. We are not used to thinking about natural languages as abstract objects. Even if philosophers such as Baker and Hacker (1984, p. 300-302) have defended that there is a mythology regarding a Platonic conception of rules, the defense of a realist (Platonic) linguistics—from a linguistic point of view—sounds very strange. However, this is exactly what Katz (1981) defends. Katz's realist linguistics presents us with an interesting proposal, which no empirical linguist has seriously considered. Some of Katz's ideas are very problematic, and present us with several challenges. Here we want to focus on his conception of natural language as an abstract object, and propose an alternative to his notion in a realist (Platonic) framework.

Katz's proposal

Katz proposes a linguistic theory whose object of study is natural language, and he conceives natural language as an abstract object. To say that natural language is an abstract object is to say that it has nothing to do with our biological constitution, or with the constitution of the concrete things we may perceive in the world. To understand this idea, we need to be able to abstract away from what we can perceive, and the comparison with mathematics, numbers, and functions simplifies this task: you know that there are numbers and functions even though you cannot touch them, you cannot see them.[3] Perhaps, by now, you are thinking that you can see numbers, that your teacher showed you how to write *2*, *2+x*, and you may think that those are numbers and functions. However, there is no actual *2*, and you cannot touch a function like *2+x*—what you may see are numerals

[1] This work was possible thanks to CNPq-Brazil.

[2] Pontifícia Universidade Católica do Rio de Janeiro. Rua Marquês de São Vicente, 225, Gávea, 22451-900, Rio de Janeiro, RJ, Brasil. E-mail: anaclarapo@gmail.com

[3] This is, at least, the default platonic conception of mathematics. On the history and problems related to this conception, see Panza and Serini (2013).

(the inscriptions of the numbers and functions). If you write them in a different typeface they will still be the same numbers, and you will still have the same function. So, we may all agree that even though we all know what the number 2 is, we can neither touch it nor see it, which is a consequence of it being an abstract object.[4] This is different to what may happen when your teacher shows you how to write *dog*. In that case, even before you knew how to write *dog*, you knew what dogs were: you saw dogs on the street, perhaps you owned one as a child. This is due to the fact that, contrary to numbers and functions, dogs are concrete objects. But what happens when your teacher tells you that *dog* is a *word?* It certainly seems to be related to those objects that you know, see, and perceive. But what is a word? This has been a topic of discussion in linguistics for a very long time, and not much agreement has been reached. This is why in linguistics the use of the term *lexical item* is preferred, because it avoids confusions involving the term *word*. However, even without taking those discussions into account, you probably are able to see the relation it may have with numbers: you cannot touch the word *dog;* you cannot see it; it seems to work exactly as *2* does. In accordance with this idea, a Platonist linguist who follows Katz would argue that words or lexical items like *dog* are abstract objects like numbers. They cannot be touched, but they are there, they are real. They are as real as a number, and we may study them. We may study words, sentences, and natural languages, which are all abstract objects.

Katz proposes a realist/Platonist linguistics because he acknowledges the abstractness of natural languages and accepts that reality is constituted by concrete objects as well as by abstract ones. These abstract objects are not created, they "have no origin and are not subject to causation" (Katz and Postal, 1991, p. 531). That is, if we accept that natural language is an abstract object we have to accept that humans did not create it, they merely discovered it, as we discovered numbers, atoms, and cells. Thus we should not confuse this abstractness with how it is usually used in linguistics: when we say that we may abstract grammar from data, for example, we are speaking about a formalization of empirical data; when Katz argues that natural language is an abstract object he is saying that it is completely independent from humans and any process of creation: "the language is a timeless, unchangeable, objective structure" (Katz, 1981, p. 22).

Katz defines linguistic theory as the study of the invariance in natural languages. That is, linguistic theory is involved in the study of linguistic universals: the features that are common to all natural languages, without which languages would not be effable (would not allow us to express every proposition by some sentence). Katz think that two examples of linguistic universals are *recursiveness* and *the principle of compositionality*. The first says that we may have very complex sentences, which are allowed by recursion. For instance, we may see language recursion in the centre embedding that allows us to embed a clause within another: *The cat that the dog chased ate the rat* (see Levinson, 2014, p. 3). The second is a reformulation of Frege's compositionality principle with a relatively small modification (in italics). The principle could, then, be rephrased as saying that "the meaning of all the infinitely many sentences and other syntactically complex constituents of a natural language *except for a finite subset of them* is a function of the meanings of their constituents and their syntactic structure" (Katz, 1981, p. 230).

While linguistic theory studies the invariance across natural languages, grammar theories are in charge of studying the specific languages and sentences of which it is composed. Grammars "correctly predict grammatical properties and relations such as well-formedness" (Katz, 1981, p. 64), and have to be able to explain grammatical facts of languages. The grammar has to be able to explain, for instance, that in natural languages there are analytic truths, such as *If John killed Mary, then Mary is dead*. In a theory like the one Katz proposes, this is explained by acknowledging that there is a link between logic and language that allows us to recognize that semantic implication may be "put on the same footing as logical implications" (Katz, 1981, p. 170). Thus, a relation between logic and language is recognized, and a more homogeneous treatment of both is defended. This approach seems to clearly explain the relations all linguists should see with logic in a very natural way: logic, language, and mathematics are all abstract objects and that's what explains the fact that we can talk, for instance, about Fregean functions and arguments in logic, in linguistics, and in mathematics.

The fact that natural languages and sentences are abstract objects, that they cannot be seen, heard, or perceived, that we may not have causal relations with them, but we still may have knowledge of them, still needs to be explained. It would seem that if we cannot see or hear them, we should not be able to speak or write. This raises the problem of access to abstract entities: how can we know them if they are causally inert, do not exist in space and time, and are inaccessible to our senses? Katz, who proposes a solid theory of linguistics, answers these questions as follows: we have a faculty of intuition that allows us to grasp abstract objects, such as numbers, sets, natural languages, and sentences. This faculty of intuition, which has to do with the knowledge we may have of a language, is understood not as sensible intuition, but as intellectual or rational intuition. He combines a realist ontology that includes abstract objects with a rationalist epistemology, and tries to provide "realism with epistemological credibility and rationalism with ontological stability" (Katz, 1998, p. xxii). Intuition, then, would be a faculty that uses an act of apprehension that allows us to know a natural language. It is through intuition that we may know abstract objects with which we have no causal relation. The faculty of intuition Katz proposes would be a general faculty, that is, it would allow us to grasp all abstract objects, not just natural

[4] Of course, this is so if you are positioned in a realist framework like the one Katz proposes. However, defining what a number is –even in a Platonist context– is not a simple task, as has been shown by Benacerraf's work: *What numbers could not be* (Benacerraf, 1965).

language: it is not language-specific, which is compatible with the fact that for a Platonist, a linguist is a mathematician, who thinks that "linguistics is a mathematical science" (Katz, 1985, p. 24). Katz recognizes that the faculty of intuition is not sufficient, that there must be other components that complement it. He proposes two extra components: one that allows us to explain how it is possible that our mental representations differ from the actual structure of natural language, and another that allows the faculty of intuition to reconstruct "representations of sentences as concrete concepts of abstract objects" (Katz, 1981, p. 205).

Katz thinks that it is very important to separate a natural language from knowledge of that language. Studying natural languages is linguistics and part of mathematics. However, studying the knowledge of language is not linguistics and is not a part of mathematics. The knowledge of natural languages would be studied, in the case we were to follow Katz, by psychology, more specifically by psycholinguistics. He argues that when we are thinking about what we produce, about what the recognition time of productions is, we are not studying the grammar of natural language but the knowledge that a speaker-hearer has of a natural language. This is due to the fact that, when we are talking about knowledge of language (competence), other factors are involved that are not part of the structure of natural language. For instance, knowledge of a language involves biological features like memory limitation, which forbids humans to produce infinitely long sentences. However, this does not mean that natural languages cannot have infinitely long sentences; it just means that humans have limitations on the knowledge they can grasp about that natural language.[5]

Katz's proposal is appealing because treating linguistics as a formal science allows us to explain at least some of the similarities that are easily seen between linguistics and logic—like semantic implication and counterfactuals, for instance. It also provides us with an argument for why grammars—in Katz's sense—almost never reflect people's real use of language: grammars study the structure of natural language and not the knowledge people have of that language. However, several issues arise here: how can we claim that natural languages are independent of humans? Why should linguistics be a part of mathematics? What does the faculty of intuition really do and how does it work? Why is the link between logic and linguistics so important? In this paper, we will focus on the question of the ontological status of natural languages, which arises from denying that linguistics is a part of mathematics.

A new definition of natural languages

The problem of the nature of natural languages arises from the idea that it is a mistake to consider linguistics a part of mathematics. Even though we agree partially with the idea that natural languages must be studied by a formal science, the linguist is not a mathematician. They share a common ground since mathematics, logic, linguistics, and computation sciences could all be seen as formal sciences that, from a realist point of view, study abstract entities. However, we think that they should be differentiated, as we differentiate the sciences that work with concrete objects (biology, chemistry, biochemistry, genetics, etc.), because they do not study the same object. Even though the faculty of intuition may be the same in grasping these abstract entities, this does not mean that they are the same entities, and nor does it mean that they should all be studied by mathematics and its "sub-sciences". The analogy with concrete physical objects comes in handy again: even though we may use perception to study an animal and to study a rock, biology and geology are different sciences. Clearly, mathematics and linguistics are strongly related, but they are not the same science. Mathematics deals with numbers and functions, while linguistics in Katz's sense deals with linguistic universals. Both are intimately related to logic, but neither of them are logic. We think, then, that Katz is right in postulating that linguistics is a formal science, but we think it should be independent of (though related to) mathematics. So, the nature of natural languages as collections of sentences has to be rethought, and a new definition of natural languages must be provided.

Whitehead and Russell (1910) proposed a hierarchy to solve the very well known paradoxes, which was created for mathematics—but may be related to linguistics. We think that we may use their hierarchy to analyse natural languages, since they are also abstract objects that are studied by a formal science. The alternative we propose allows for a better definition of the ontological status of natural languages. It is based in a logic-ontological hierarchy like the one given by Chateaubriand (2001)—inspired by Whitehead and Russell (1910)—that includes objects, logical and non-logical properties, and states of affairs, and is a flexible and cumulative hierarchy. In our hierarchy, at level 0 we have concrete entities (objects and events), and from level 1 up we have abstract entities, such as properties, facts, and states of affairs. Both facts and states of affairs are complex entities that result from the instantiation of a universal property in a concrete entity/ies or in an abstract entity/ies, respectively. We have non-logical and logical properties. Non-logical properties appear at level 1, and logical properties may appear at different levels and combine with different types of entities, which makes our hierarchy flexible and cumulative.

We would like to propose that, in this hierarchized ontology created for mathematics and logics, it is possible to have linguistic abstract entities that may also be defined according to the place they occupy in the hierarchy. They belong to the same ontology, for even though they should be studied by dif-

[5] See Langendoen and Postal (1984) for a proposal of natural languages that have infinitely long sentences.

ferent sciences, they are abstract entities. As abstract entities they may be properties or states of affairs, and appear in level 1 of the hierarchy for the first time. Linguistic properties will be all the linguistic items that may be combined with other linguistic items (lexical items, grammatical rules, meaning, etc.). Linguistic states of affairs will be the result of the instantiation of linguistic properties in other linguistic properties (for instance, the instantiation of the property *being spoken American English*, in the lexical item *cat*). The hierarchy, thus, results in a more complex and rich linguistic ontology than the one proposed by Katz. However, if the ontology only had sentences and collections, and natural languages were defined as collections of sentences (following Katz), many linguistic phenomena that involve smaller linguistic units than sentences and "combinations" of collections would not be explained. Our hierarchized ontology, in contrast, allows us to explain diverse linguistic phenomena, such as linguistic change, the existence of dialects, sociolects, and linguistic family trees, among others.

But, before explaining those phenomena, we should decide which linguistic entities appear at level 1 (which is not an easy task). On the one hand, it would seem that the phonological level should be the first, since we could not have lexical items without phonological structure. Phonological rules and phonemes could be treated as abstract entities. They may be grasped, someone may utter /o/, which means that they accessed the type of the phoneme and transformed it into a token. But do phonemes and phonological rules belong to the same level, or do phonological restrictions appear first, then phonemes form on a higher level, and then combine? This is not an easy question to answer, and more needs to be known about the nature of phonology before we can stipulate any answer. On the other hand, it seems that meaning should be prior. In fact, it seems that meaning should be universal (common to all natural languages), and thus everything linguistic—even phonemes—should appear on higher levels than that of linguistic meaning. They could be considered linguistic primitives, and appear at level 1. Then, every other linguistic entity would be related to that meaning in a direct or indirect way, and exist from level 2 and up. However, this is not an easy problem to solve, because we should explain how we relate meaning to everything that is in the ontology—which is not a task we undertake in this paper.

Another important feature that should be taken into account when trying to build a linguistic ontology is morphology. Morphology is complicated, and defining where morphemes stand is an even more complicated task. We would have to decide which is the best approach to morphology. Do we first have morphemes and morphological rules that form lexical items, and then go on to form sentences (which would seem similar to the 'Weak Lexicalist Hypothesis,' like that proposed by Aronoff, 1976); or do we have all lexical items (including morphemes) on the same level and all the grammatical rules on a higher level which, once instantiated, have as a result lexical items, phrases, and sentences (which would be something like the 'Distributed Morhpology' framework, see Harley and Noyer, 1999)? Obviously, an answer to this question should be given, and it seems that the grammatical rules used to form words and sentences are very similar, which should point to option number 2—but this should also be further studied.

Discovering which level natural languages appear on for the first time is not an easy task, and is not one we can solve here. This is due to the fact that many linguistic problems are still not solved. Nevertheless, this does not mean that we are unable to explain how natural languages could be represented in a linguistic ontology like the one we propose. To do this, we may simplify our ontology, start with lexical items at level x, and accept that x is an undefined level that depends on how many levels turn out to come before the lexical items of a particular natural language are formed. We leave for future research, then, the levels that are yet to be completely discovered, and start at x.

Starting at x

Let's start at level x, because we still cannot say which level lexical items are found on. We may still consider lexical items as independent entities that can be combined with other lexical items. That is, lexical items could be treated as linguistic properties that are autonomous, and able to combine with each other through grammatical rules. Since they are the most basic and simple abstract linguistic entities that may form phrases and sentences, they pertain to level x. Lexical items have meaning and phonological structure (/kat/, for instance, would be a possible representation of a lexical item, and *cat* a simpler one). Then we would have grammatical rules. Grammatical rules are properties that establish how those lexical items may be combined to form phrases and sentences ('a transitive verb combines a subject with a direct object,' for instance, could be a possible representation of a grammatical rule), and, thus, should pertain to higher levels than that of lexical items. They should appear for the first time at x + 1. Phrases and sentences, then, result from the combination of lexical items and grammatical rules. We propose that phrases may be interpreted as the instantiation of a grammatical rule in lexical items (by analogy with states of affairs), and thus they must belong to the level x + 1—the same level as the grammatical rules. A level x + 1 phrase like *the cat* would be the instantiation of a grammatical rule in two lexical items *the* and *cat*. Sentences would result from the combination of level x + 2 rules and level x + 1 phrases. A level x + 2 sentence like *Paula loves the cat*, for instance, would be the result of the instantiation of several grammatical rules in several phrases.[6]

[6] By assuming this, we are assuming that lexical items alone cannot combine to form a sentence. So, even though, *Paula* may appear to be only a lexical item, it is a DP.

Once we have sentences, we may have natural languages as level x + 3 entities (or linguistic states of affairs) that result from the instantiation of a property like *being spoken American English* in lower level abstract entities. This level x + 3 property has a variable arity, and may be instantiated in a lower-level property (lexical items or grammatical rules) or state of affairs (phrases or sentences). However, it will be a natural language only in case it is instantiated in all the entities that pertain to spoken American English.[7] This is due to the fact that we want it to be possible for a level x + 3 property like *being spoken American English* to be instantiated, for instance, in only one sentence, as in <*being spoken American English, Paula loves Pedro*>, as well as in all the sentences, all the grammatical rules, all the lexical items, and all the linguistic items that pertain to spoken American English. We can thus explain how we differentiate what belongs to English from what belongs to Spanish: only those entities in which *being spoken American English* is instantiated pertain to American English, and only those in which *ser español rioplatense hablado* is instantiated pertain to that variety of Spanish. This allows us to explain why we have *dialects* that pertain to a natural language. We could state that those level x + 3 states of affairs are Dialects of a level x + 4 Standard Variety, which would result from the instantiation of a property like *being the Standard Variety of American English* in lower level entities, which would itself account for standard languages of a given variety. That is, the standard language is the result of the combination of a level x + 4 property with lower level entities that explain how we can have 'one' American English.

Once we have those standard level x + 4 natural languages (NL), we may then have a level x + 5 NL which contains the common structure of all those standard level x + 4 NL that may be said to be English. In this case, we would have a property like *being the Natural Language English* that would be instantiated in lower level entities, and which would account for what is usually called English (or Spanish, or Chinese, etc.). This would explain why, even if it seems obvious that there is not a unique English, we may speak of English as being one language: we talk about the common structure of different standard varieties of English (or any other language). We then would go up to level x + 6, in which the properties of *being a Family of Natural Languages* (FNL) appear for the first time. In this case, the property would be instantiated in level x + 5 Natural Languages, resulting in the first branch of a linguistic family tree. After level x + 6 Family of NL, we would have a level x + 7 Family of NL that contains level x +6 FNL; and so on, until everything has been explained. This is an exciting idea that allows us to explain not only dialects and standard varieties, but also linguistic family trees: when we are doing typology, for instance, we are discovering those families of NL that belong to a higher level, say level x + 6, and to which level x +5 NL belong.

We will illuminate our point with a rather simplified linguistic family tree of the Germanic languages—specifically the English branch (for further information see Mallory and Adams, 2006, Ch. 2.4). First, we start with level x + 3 NL which are those that result from the instantiation of a level x + 3 property in lower level entities (which may be level x + 2 sentences, or level x lexical items, etc.) and were spoken at some time but are not the standard varieties considered as level x + 4 NL. Then, in the case of the English branch, we would have the level x + 4 standard varieties of Old English (spoken from 700–1100), of Middle English (spoken from 1100–1450), and of New English (spoken from 1450 until now, which should be further divided into the standard varieties of the dialects currently spoken in England, and other English-speaking countries). The standard varieties, level x + 4 NL, are not spoken by most members of the communities. They are studied by grammarians, they may be used by illustrated members of the community, but—since they include some items and disregard others—they may only be used by those that have studied the grammars of the varieties. For instance, a speech given by an outstanding member of English society would most probably be given in standard British English. After level x + 4, we would go up to level x + 5 NL, which contains level x + 4 NL. They have a common linguistic structure that allows us to classify them as English. Since level x + 5 NL contains natural languages of natural languages (and include their common structure), even though they are apprehended, they are not spoken. They might be studied by grammarians and compared to other level x + 5 NLs, but they are not spoken by communities (nor by those that may speak the standard varieties). After level x + 5 English—which includes the common structure present in all level x + 4 English standard varieties—, we have a level x + 6 family of Natural Languages, which is Germanic, and which is composed of the following level x + 5 NL: English, Dutch, West Germanic, North Germanic, and East Germanic (which are constituted in a similar way to English). This is a relatively simple example, if compared to a whole tree, of how linguistic family trees may be explained by a proposal like ours. It could be followed by a Proto-Indo-European Family, which would be a higher-level natural language (as a linguistic state of affairs) that contains all the NL that are a part of it, but that would be better developed once we know exactly how the levels up to x + n work (for other groups that pertain to Indo-European and Proto-Indo-Eruopean, see Mallory and Adams, 2006, table 1.6, p. 9).

Concluding remarks

We think that our proposal goes hand in hand with Katz's. It maintains that Natural Languages are abstract entities, and that linguistics is a formal science, though not a part of mathematics. The characterization of NL according to the different levels of the hierarchized ontology presented seems

[7] We are using "spoken English", because the level x + 3 natural languages would be the actual spoken varieties.

to be a nice addition to Katz's account because it allows us to explain how languages are related. It also allows us to avoid taking NL as abstract objects. Treating NLs as states of affairs, as combinations of properties, allows us to explain NLs as complex entities that involve relations between different linguistic entities. On the first level analysed, x, we have lexical items; on the second level analysed, x +1, we have grammatical rules which, combined with lexical items, result in level x + 1 phrases; on the third, x + 2, we have grammatical rules and sentences; on the fourth, x + 3, NLs (subdivided in dialects, standard varieties and NLs) appear; we then go up to the level on which families of NLs appear, but NLs do not appear on lower levels. This allows us to explain many linguistic phenomena that could not be explained if NLs were objects, such as collections of sentences. This notion of NLs obviously involves some problems that should be further investigated, as mentioned, but we are convinced that they are worth researching.

References

ARONOFF, M. 1976. *Word Formation in Generative Grammar*. Cambridge, MIT Press, 134 p.

BAKER, G.P.; HACKER, P.M.S. 1984. *Language, Sense and Nonsense. Critical Investigation into Modern Theories of Language*. Oxford, Basil Blackwell, 397 p.

BENACERRAF, P. 1965. What Numbers Could Not Be. *The Philosophical Review*, **74**(1):47-73. https://doi.org/10.2307/2183530

CHATEAUBRIAND, O. 2001. *Logical Forms, Part 1*. Campinas, Coleçao CLE, 442 p.

HARLEY, H.; NOYER, R. 1999. State-of-the-Article, Distributed Morphology. *Glot International*, **4**:3-9.

KATZ, J. 1981. *Language and Other Abstract Objects*. New Jersey, Rowman and Littlefield, 251 p.

KATZ, J. 1985. An Outline of Platonist Grammar. *In:* J. KATZ (ed.), *The Philosophy of Linguistics*. Oxford, Oxford University Press, 266 p.

KATZ, J. 1998. *Realistic Rationalism*. Cambridge, MIT Press, 262 p.

KATZ, J.; POSTAL, P. 1991. Realism vs. Conceptualism in Linguistics. *Linguistics and Philosophy*, **14**:515-554. https://doi.org/10.1007/BF00632596

LANGENDOEN, T.; POSTAL, P. 1984. *The Vastness of Natural Languages*. Oxford, Basil Blackwell, 198 p.

LEVINSON, S. 2014. Pragmatics as the Origin of Recursion. *In:* F. LOWENTHAL; L. LEFEVBRE (eds.), *Language and Recursion*. New York, Springer, p. 3-13. https://doi.org/10.1007/978-1-4614-9414-0_1

MALLORY, J.P.; ADAMS, D.Q. 2006. *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. New York, Oxford University Press, 731 p.

PANZA, M.; SERINI, A. 2013. *Plato's Problem: An Introduction to Mathematical Platonism*. London/New York, Palgrave MacMillan, 306 p. https://doi.org/10.1057/9781137298133

WHITEHEAD, A.N.; RUSSELL, B. 1910. *Principia Mathematica*. Cambridge, Cambridge University Press, 658 p.

Submitted on November 02, 2016

Accepted on March 13, 2017