TEACHING LINGUISTIC METHODS WITH CONLANGS

Guillaume ENGUEHARD

ABSTRACT • In this article, I show from my own experience as a linguistics teacher that the construction of languages by students allows them to better grasp the reality of linguistic analysis than does the repetition of exercises based on data from natural languages. I show that the reasons that lead me to marginalize natural languages from teaching are the same ones that lead most linguists to marginalize constructed languages from research.

KEYWORDS • Natural Languages; Natlang; Linguistic Analysis; Constructed Languages; Conlangs; Diachronic Linguistics; Linguistic Pedagogy; Method-Oriented Teaching.

1. Introduction

I warn my reader that this article does not represent a scientific approach to didactic methods in linguistics. Rather, I present a teacher's perspective on what is taught in linguistics and how it is taught. My contribution is to point out the difference between the subject, the object and the method in the field of study, and to argue that the traditional subject-oriented teaching paradoxically addresses more artificial data than method-oriented teaching. I will base my proposition on a reflection from my last years of teaching phonology and diachronic linguistics at the undergraduate level. At no point is my goal to show whether one or the other of these two methods leads to better results. While I have my opinion on this and I see greater student enthusiasm for a method-oriented approach, I have no way of measuring this contrast objectively or guaranteeing that these results are reproducible with any audience. This depends on many factors such as the students' professional project, their cultural background, the teacher, etc. Therefore, this article should only be understood as a critical analysis of what makes the objective of teaching linguistics to undergraduate students.

Generally, linguistics is considered as a discipline that aims to study natural languages. This is why constructed languages (henceforth 'conlangs') are often at the margin, if not absent, from the teaching of linguistics. They are not considered as genuine subjects of the discipline. Whether this proposition is accepted or not, it is nonetheless true that conlangs mostly mimic the structure of natural languages. Therefore, they can be analyzed using the same protocols as those applied to natural languages. These protocols of analysis are what I will call the 'method' in the first section.

Usually, the linguistic method is not taught separately from the linguistic subject. The exercises proposed to the students generally deals with analysis tasks applied to small data from natural languages. This is what I call a subject-oriented teaching: students are introduced to

linguistic methods via the study of natural languages. I argue in the second section that the data extracted from these languages are not so natural.

In the third section, I show that the use of constructed data ironically allows students to put themselves in a situation closer to the reality of the fieldwork analysis. This is what I call methodoriented teaching: students are introduced to the subject of linguistics via the use of linguistic methods.

1. Subject and method

In this section, I introduce three aspects of any field of study: the subject, the method and the object.

1.1. Subject

I call a 'subject' any set of facts from which questions emerge and which justify the definition of a field of study. To this respect, languages are a subject. It is by observing languages that we come up with questions concerning their relationships, their internal organization, their use, etc. All these questions are gathered in a field of study called linguistics.

There are many types of languages. Among them, we find natural languages, revived languages, conlangs, etc. But are all these languages a subject of linguistics? Technically speaking, no. Unlike natural languages, conlangs generally *result* from a linguistic questioning. For instance: Zamenhof (1887) sought to gather the most common structures of European languages in his Esperanto (Waringhien, 1967); Lang (2014) tried to maximally reduce the complexity of language in her Toki Pona; and even Tolkien focused on the etymology of Old English *éarendel* in his Elvish languages (Tolkien, 1981, p. 385-386). Because they are posterior or at best concomitant to the linguistic questioning, the conlangs cannot be defined as primitive subjects of the field of study.

This does not mean that it is impossible for a conlang to be a subject of linguistic questioning, but that it is not primarily so. It can become so secondarily, provided that it becomes independent of any individual participating in its construction. This may be the case of only some conlangs such as Esperanto. This does not mean either that the remaining conlangs cannot or do not deserve to be studied, as we will see later.

1.2. Method

I call 'method' a set of rules by which one proposes to answer the questions that form a field of study. These rules form what is called a scientific discipline. The term linguistics is equivocal, since it designates both a field of study and a scientific discipline governed by its own methods.

Linguistic methods are numerous, but they all aim at analyzing a subject, i.e. a language, and to extract its structural essence. I take here two examples in connection with the contents of the courses that I will present later. First, the phonological method consists in defining the relevant units that form the signifier of words by noting the oppositions of form and meaning and by ignoring all the contrasts that are not functional. Second, the diachronic method consists in comparing one by one the functional units of words from different languages in order to find regular equivalences that reveal a common origin. We could thus summarize at length all the fundamental methods of structural linguistics, covering semantics, morphology, syntax, etc.

These methods are logically applied to natural languages, in order to understand their functioning. But, more importantly, they can be applied equally to conlangs. Indeed, whatever their stated purpose, all conlangs have in common that they imitate the functioning of a language, either by copying the structure of a natural language or by starting from universal principles such as communicative function and energy saving. Thus, it is not surprising that one can produce the linguistic analysis of a conlang. It is the case with auxiliary languages (de Saussure, 1914; Kaloczay and Waringhien, 1935) and also more artistic languages (Salo, 2004).

However, the question is not so much whether it is possible to apply linguistic analysis methods to conlangs but rather what the value is. As we have seen, conlangs do not precede linguistic questioning. They follow it as a response. Take the example of the Brithenig created by Andrew Smith.¹ First comes the question of the hypothetical appearance of Latin after several centuries of evolution in the British Isles. Then comes the method that allows us to answer this question by deriving the Brithenig language. Analyzing this language afterwards does not allow to make more discoveries with regard to the initial questioning, since the simple existence of Brithenig is in itself an exhaustive answer. Therefore, at first glance, it seems unnecessary to apply linguistic methods to conlangs.

1.3. Object

What distinguishes a conlang from a natural language with respect to linguistic research is the fact that a conlang, by its very essence, is the application of answers to the question that the linguistic method seeks to resolve. Therefore, it is not a subject of linguistics, but what I will call an 'object' of linguistics, that is to say a production implementing the answers that were found to the linguistic questioning via a certain method. I will call the production of such an object an application of linguistics. The concepts I have introduced so far are summarized in the diagram below. Questioning is the induction of a method from a subject. Application is a deduction of an object from a method. In this sense, an object of linguistics can be a conlang, a theory, a protocol or anything else that follows from the method.



Figure 1: The research path in linguistics

¹ http://steen.free.fr/brithenig/introduction.html.

The comparison between a conlang and a theory is particularly relevant, because it allows us to realize that the subject and the object of a scientific discipline must necessarily be distinct. Otherwise, we could imagine that a theory is both the object and the subject of the discipline, which would lead to a certain form of circularity. The same is true with conlangs. A linguistic discipline that defines a questioning from conlangs and formulates its answers with the same conlangs would end up becoming circular. Thus, conlangs may be the subject of linguistics only secondarily.

This is why, in the examples I cited above, a conlang generally becomes the subject of linguistics only when it escapes its creator, either because its creator no longer intervenes in its development, or because it becomes a collective phenomenon. This is the case of Esperanto, which has become a language developed by use although it is of constructed origin. But this is also the case of Tolkien's Elvish languages, which are incomplete and whose gaps can only be filled today by their own internal consistency.

To conclude this first section, I have shown that although it is sometimes legitimate to focus the subject of linguistics on conlangs, this can only be done secondarily and in very particular cases. I will now show that, paradoxically, the teaching of linguistics via natural languages suffers from the same defects mentioned just now.

2. Subject-oriented teaching

I define here the principles and shortcomings of a classical, subject-oriented teaching of linguistics.

2.1. Main principles

The classical, subject-oriented, teaching of linguistics consists in introducing simultaneously the knowledge of the methodology and the knowledge of the subject. In concrete terms, students are led to understand the methodological principles through the example of natural languages, and to apply the same methodology to data from other natural languages. The following table is an exercise, taken from the textbook by Kenstowicz and Kisseberth (1979), that illustrates this method. Given this data from Russian, students are asked to look for a context of neutralization.

Nom. sg.	Dat. sg.	Gen. pl.	Gloss
ryba	rybe	ryp	`fish`
tropa	trope	trop	'path'
pobeda	pobede	pobet	'victory'
sirota	sirote	sirot	'orphan'
groza	groze	gros	`storm`
krysa	kryse	krys	'rat'
lyža	lyže	lyš	`ski`
duša	duše	duš	'soul'
noga	noge	nok	'leg'
sobaka	sobake	sobak	'dog'

Figure 2: Example of exercise (Kenstowicz and Kisseberth, 1979)

CrOCEVIA · Conlangs and Linguistics: Theories, Practices, Case Studies

At first glance, some would say that this is what linguists do and that this method of teaching therefore prepares students for the tasks that await them in a real situation. But this is not the case. This method has formal and substantive problems that I present below.

2.2. Shortcomings

2.2.1. Limited data

On a strictly practical level, subject-oriented teaching is contingent on the availability of data from natural languages. Of course, one can assume that any good teacher of linguistics has a sufficient database for the construction of examples, exercises and test materials. However, this database is not infinite and its content always ends up going in circles. This limitation of data poses a problem when one encounters a student who fails from one year to the next and therefore ends up following the same examples and exams.

Second, this is a constraint that also weighs on the teacher's practice. The latter must adapt to the available data and not the other way around. For example, it is sometimes difficult, if not impossible, to find several exam subjects of equivalent level and based on large databases. This is especially true in diachronic linguistics, where students must reconstruct prehistoric roots from a comparative analysis. Finding a large number of related lexical items in a language family and obtaining their phonetic transcription is, except for a few academic examples, mission impossible. It is a work that comes under the heading of linguistic description rather than didactics, and is therefore part of a much longer time frame. The teacher must therefore limit the number of exercises and exams that he/she offers to the students. He must often use the same subject for all, which increases the possibilities of fraud.

2.2.2. Unrepresentative data

More importantly, course data is often unrepresentative of the reality of languages or the reality of language research. First, the exercises often reflect the research interests of the teacher. A teacher working on African languages will present more data from African languages, a teacher working on European languages will build his or her exercises on the basis of English, French, or Russian, a teacher working on tonal languages will present phenomena related to tones, etc. Thus students are subject to either areal or disciplinary bias, and their intellectual maturation therefore depends on uncontrolled factors.

This first aspect may seem unavoidable, and it may be said that a student's learning is an extension of the teacher's knowledge. But this becomes problematic when the teacher's knowledge ends up representing blinders for the student. Indeed, by presenting examples of known phenomena, the teacher presupposes what we expect to find in languages. Presenting numerous examples of voicing assimilation and intervocalic lenition allows students to assimilate these concepts, but does not allow them to open up to rarer phenomena.

This is the key issue. The data presented in a teaching context are not real data, but the result of descriptions and analyses. They are often provided in an ordered form and therefore do not reflect the chaos that the linguist is first confronted with. In the Russian example I cited above, the data are not only ordered according to their morphological function, but also according to their phonological properties. Each consonant is represented by a line, and the lines follow each other according to the place of articulation of the consonant and its voicing. No field linguist encounters such an idyllic situation in real life. With such data, students are ultimately only able to describe data whose presentation presupposes description, but they are generally not able to find order in the chaos. This brings us back to the problem of limitation. If the teacher cannot present raw data, it is often because he or she only has access to published data from language analyses, not real data.

In the first section, I had shown that the constructed languages, because they are objects of linguistics, could not represent a primary subject of the latter. Otherwise research would only go round in circles. This situation is comparable to that of analysed data in linguistic teaching. Since the analyzed data are the result of the linguistic method, i.e. its object, they cannot serve as the subject of the latter without leading to a circular reasoning.



Figure 3: Analyzed data are an object of linguistics

This is a paradox, because for practical and theoretical reasons, subject-oriented teaching is also inevitably object-oriented.

3. Method-oriented teaching

In this section, I show that language construction in linguistic teaching avoids the aforementioned shortcomings by focusing exclusively on methodological principles.

3.1. Main principles

The construction of linguistic data in the teaching of the discipline is inspired by active pedagogy in social sciences (Cavanagh, 1975; Laveault and Corbeil, 1986; Chambelland and Provost, 1996) as well as mathematical and physical problems, which often involve imagined situations, often absurd, from which students must learn to apply a certain number of acquired principles. As a teacher in the department of linguistics of the University of Orléans (France), I applied this method in classes of 25 to 35 undergraduate students working on phonology or diachronic linguistics from the year 2020 to the year 2022.

In the field of history teaching, Cavanagh (1975) shows that simulation games do not distort reality and, on the contrary, allow students to go beyond the limits of historical reports by becoming aware of the historical process itself, the simultaneity of its factors and the diversity of viewpoints (see also Laveault and Corbeil, 1986).

In practice, applied to linguistics, this method consists of asking students to create languages from scratch and then produce an analysis of them. Of course, to avoid the situation in which the subject and the object of the method become confused, the students are given limited freedom as to the structures of their languages. Both in phonology and in diachrony, the first elements of the lexicon arbitrarily defined by the students go through the milling of phonetic change rules which bring to the language a certain number of phonotactic constraints which distinguish it from its root.

The example below is an excerpt from a swadesh list showing two dialects derived with random change rules from a primary language created by a group of five students in a phonology class in 2021. As you can see, varieties A and B have undergone changes such that the vowels [æ] and [u], as well as the consonants [m] and [ʁ], are in free variation.

	Dialect A	Dialect B
to	['wæz]	['wuz]
wing	['āvæz]	['āvuz]
year	['izæm]	'izus
animal	['duv ^j iz]	['dæv ^j iz]
tree	[wõv]	[wov]
other	['edæm]	['edus]
with	['J.)Z]	['G.)7]
pole	['vat]]	['vat]]
many	['im ^j i]	'isi
white	['3 ^j e]	['3 ^j e]

Table 1: Example of dialectal variation generated by a group of students

In diachronic linguistics, this step is repeated several times by smaller and smaller groups of students, along with random semantic shifts and synonymic replacements, so as to obtain a family of related but sometimes very distinct languages. To these change rules are added the (very) numerous errors of the students which bring to the language a part of chaos simulating for example analogy effects and borrowings between dialects. Table 2 (p. 10) shows some members of a language family derived from the same parent language by an entire class of students in 2020. Capital letters represent the names of the languages.

As you can see, some words are identical (e.g. I and J words for '*wing*'), some are similar enough (e.g. F and J words for '*other*', and some have obviously undergone a synonymic replacement (e.g. D and E word for '*tree*'). The gaps are due to students not having completed their work, but they do not disturb the process of the exercise as they can correspond to the lack of accurate translation between two structurally different languages in real life.

In short, students only have control over the source of the teaching material, not its final state. Conversely, the teacher has some control over the changes undergone by these languages, but not over their initial and final structures. Unlike the data of the subject-oriented teaching method, the data resulting from this process is not controlled by any of the actors of the class and therefore fairly faithfully reflects the reality of what the field linguist is faced with.

Table 2: Example of related words generated by students

CrOCEVIA · Conlangs and Linguistics: Theories, Practices, Case Studies

3.2. Analyzing real data

This is the first advantage of this method. The data thus produced represents a substantial volume, easily accessible and possibly reusable from one year to the next (although it is more interesting to generate new data every year). The teacher is therefore no longer limited in his practice. In addition, this raw data gathers items without ordered presentation. Students must therefore learn to observe the data carefully in order to be able to spot regular elements.

Figure 2 for example provides a table that has been realized in 2020 by a student in diachronic linguistics looking for concurrent equivalence rules in a raw database. The first four columns represent different varieties derived from the same proto-language. The first sub-table show three examples of an equivalence between the [u] vowels of related words, and the following sub-tables show concurrent equivalences between the same languages.

Chouo'a	Chouota	Jouar	Tchou	Glose	Règle
puʒurvu	puhurvu	puurvou	pulho	avariée	ulululu
funjurpp	funjurpp	vunjur¢pp	funjulbo	insuffisance veineuse	ulululu
k ^h u	gu	gu	gu	tout	ulululu
k ^h u	gu	xou	go	creuser	ululolo
tu	tu	SO	so	grand	ululolo
ru	ru	rou	lo	os	u u o o
t ^h u	su	zuu	su	condiment	u u uu u
ru	rur	ruu	lu	laver	ululuulu
knu	knu	guu	ku	vous (pluriel)	u u uu u
k ^h u	gu	xou	go	creuser	Ø Ø u Ø
ru	ru	rou	lo	OS	ØØUØ
kuo	ku	gu	ko	voler (dans l'air)	ØØUØ

Figure 4: Example of concurrent equivalence rules defined by a student

More than just in the presentation of data, this chaos is also found in the data itself. Thanks to students' errors, a phenomenon (a phonotactic constraint or a phonetic correspondence) can be regular but not absolute. Thus, students do not get used to having data that is too clean and are led to think about the very notion of regularity.

Better still, depending on the starting point of the constructed language, a phenomenon may be completely absent. The students (like the teacher) therefore know neither what they will find by observing the data, nor even if they will find anything there. This is an essential element that forces them to be more autonomous and to overcome their teacher's knowledge. It is important to specify that they always manage to find something, even if it is sometimes after long efforts and a few attempts to give up.

Finally, and this is the most important, these three points lead the students to adapt the method to the reality of linguistic description. Classical teaching often reinforces them in a strict application that has no use outside the class. The most telling example is that of phonological oppositions. Telling

students that an opposition must be based on a minimal pair leads them to think that an absence of a minimal pair implies an absence of opposition. Unfortunately for us, the reality is often much more complex and less formal. Oppositions can sometimes only be argued on the basis of quasi-minimal pairs, the absence of free or contextual variation, and the symmetry of the phonological system.

In short, with constructed languages, students face the reality of raw data and linguistic description. They learn the difference between the method, founded on a reasoned basis, and its concrete application.

3.3. Connecting dynamics and structure

These advantages are enough to circumvent the defects of the classical teaching method. But there are other advantages to using language construction. The first of these is that which consists in leading the students to question themselves on the link between diachrony and synchrony. As I have described, students construct the data they analyze from change rules provided to them. It is these rules, among others, that define the free variations, contextual variations, neutralizations and phonetic equivalences that they can then observe. Therefore, the relationship that exists between the dynamics of a language and the resulting structure is explicitly shown to them.

However, it is clear that most students have difficulty establishing this link. Thus, after having spent one or more sessions changing the intervocalic obstruents into voiced obstruents, few of them can deduce a constraint prohibiting the presence of voiceless obstruents in the intervocalic position. These students can only find a constraint by direct observation.

Conversely, other student profiles with a more analytical mind content themselves with a change rule to define a complementary distribution or a neutralization. But that's without taking into account the initial state of the language, other rules and the errors of application. These deductions sometimes turn out to be wrong in practice and the students then realize that the structure of the language can result from contradictory changes, and that a deduced generalization must always be validated by direct observation of the data.

3.4. Comparing truth and validity

This access to both the dynamics and the structure of the language allows a more general reflection on the difference between the validity of a reasoning and the reality of the observed phenomenon.

In diachronic linguistics in particular, students have the opportunity to compare the roots they reconstruct by applying the method, with the real roots from which they make their languages evolve. Apart from a few examples that always bring us back to the limitation of data (e.g. Romance languages and Latin), this comparison is not possible with natural languages. Sometimes, even when the method is correctly applied, the roots encountered do not exactly correspond to the real roots. The students must then produce a reflection on the reasons for this contrast and thus understand, for example, that the method does not make it possible to reconstruct an element that has disappeared in all the daughter languages. In the following translated excerpt dating from 2020, a student in diachronic linguistics tries to explain the contrast between the reconstructed root and the real root by questioning whether this contrast is phonological or not.

In conclusion, the reconstructed form of the Pre-Proto-Chuan nominal affix is ***-bb**. After verification, the earliest attested form is Eastern Chuochaar **-pbb**. The reconstructed root is therefore not factually true. **The problem is that we observe that [p] and [b] are in opposition in this language: presence**

of a minimal pair (bob '*stick*' pbb '*to*'). The deviation from the reconstructed form is therefore linguistically relevant. We had determined that p|b was derived from *b, but we may have been wrong. Indeed, in the absence of fusion or fission (since the alternation is isolated), we found it difficult to arrive at a relevant reasoning to really know if p|b was issued from *p or from *b. It is therefore likely in the end that p|b is from *p. In this case, the reconstructed form of the Pre-Proto- Chuan nominal affix would be *-pp, which would not be too far from -ppp (in Eastern Chuochaar). Indeed, we observe that the [p] (short) does not exist in this language, we only observe a [pp] (long). The contrast with the reconstructed form is therefore no longer linguistically relevant.

An anonymous student

It is necessary to point out that not all students produce such a sophisticated reasoning.

Conclusion

To conclude, I showed that the construction of languages during a linguistics course represents an advantage compared to the use of natural languages. After first explaining the reasons that intuitively lead linguists to exclude constructed languages from their teaching practice, I have shown that the use of published linguistic data leads these same linguists to do exactly what they seek to avoid. Conversely, the use of constructed languages allows students to face the reality of linguistic analysis and to question the limits of the method.

The limit of this new method-oriented teaching is that it does not alone allow students to acquire knowledge of natural languages. It is therefore important to combine it with a more theoretical teaching introducing the methodological principles by illustrating them with examples from natural languages.

REFERENCES

Cavanagh, T. K. (1975). *Simulations Gaming in Canadian History*. Sherbrooke, Publications Progressives. Chamberland, G. and Provost, G. (1996). *Jeu, simulation et jeu de rôle*. Sainte-Foy, Presses de l'Université du Québec.

- de Saussure, R. (1914). La vort-teorio en Esperanto. Universala Esperantia Librejo.
- Kaloczay, K. and Waringhien, G. (1935). *Plena analiza grammatiko de esperanto*. Universala Esperanto-Asocio, Rotterdam, [1985] edition.
- Kenstowicz, M. and Kisseberth, C. (1979). *Generative Phonology: Description and Theory*. Academic Press, New York.

Lang, S. (2014). Toki Pona: The Language of Good. CreateSpace, Charleston.

- Laveault, D. and Corbeil, P. (1986). Psychopédagogie du jeu de simulation pour l'apprentissage de l'histoire. *Revue des sciences de l'éducation*, 12 (1), 25-43.
- Salo, D. (2004). A Gateway to Sindarin: A Grammar of an Elvish language from J. R. R. Tolkien. The University of Utah Press, Salt Lake City.
- Tolkien, J. R. R. (1981). The Letters of J. R. R. Tolkien. Houghton Mifflin, Boston.

Waringhien, G. (1967). ABC d'Esperanto. SAT.-Amikaro & U.F.E., Paris.

Zamenhof, L.-L. (1887). Mezhdunarodnyj Jazyk. Kh. Kel'ter, Varsovia.

GUILLAUME ENGUEHARD • is Assistant Professor of Linguistics at Université d'Orléans and the research lab LLL (UMR 7270 CNRS); he works in the field of theoretical phonology and has published articles on the role of time units and phonological contrast in stress and syllable structure. He is also working on defining the varieties of Old Norse formerly spoken in Normandy.