On Syntactic Formalism(s) and Ancient Indo-European Languages

Federico Giusfredi

-

federico.giusfredi@univr.it

MSCA Horizon 2020 Project SLUW

University of Verona

Part 1: Frameworks

What is a framework?

Haspelmath 2008: «a sophisticated and complex metalanguage for linguistic description that is intended to work for any language»

Problems:

- it is not necessarily "complex" (besides, "complex" is a very ill-defined label);
- not every part of a framework need to be universal (a framework can claim to be cross-linguistically complete, but it must not necessarily do so).

Better definition: a **set of pre-constructed conceptual tools (mappable over a metalanguage). A complying scientific description is expected to fit the framework *or to modify it*.**

Frameworks exist in other sciences, too (e.g. Special Relativity was a partial framework of physics, and it was expanded and generalized by the later development of General Relativity).

In general linguistics

Several frameworks existed even before the word “framework” became fashionable. Consider e.g.:

- Historical linguistics (genealogical vs. wave models in the 19th century);
- Early typology (à la N. Marr);
- Structuralism (F. De Saussure, Copenhagen School, Geneva School, Prague School...);
    - Functionalism (from Martinet to Dik, Halliday, Van Valin...)
- Typology (Greenberg...)

- ...and so on... (for an updated critical discussion of modern linguistics - and linguistic concepts and schools -, cf. Graffi, 2006 & 2013)

In historical linguistics

Neo-Grammarian Historical linguistics (and Indo-European linguistics as a part of it) **is a framework itself**. It dictates, for instance: genealogical filiation, monogenesis of the Ursprache, taxativity of sound change under identical conditions. As such, it includes laws, that can even be formalized. For instance, one may give the following formal definitions that are assumed to hold for every language family and for every sound change:

Def. Genealogical filiation: Let L' be a language. $\forall$ L' $\exists$ L | L > L' (there is always a mother-language)

Def. Monogenesis: Let U be a set of languages L(i). $\exists!$ L(u) $\notin$ U | $\forall$ L $\in$ U L(u) > L (the mother-language of a finite set of languages is unique)

Def. Taxativiy of sound change (on phonemics and context-sensitiviy cf. Chomsly/Harris, 1968, The sound pattern of English; also, on the computation of synchronic phonological rules, Kaplan/Kay 1994, 'Regular Models of Phonological Rule Systems', in JCL 20/3, pp. 331-378):

Let X be a non terminal element within a formal-grammar-generated language, z and y be series of non-null elements in a string S = zXy. Let the language L be the derivational non-terminal language that includes zXy. Then a sound-change from L to L' has the form $\forall$ zXy in L zXy > zWy | W $\in$ L'.

NB: these principles are still valid, but their universality has been challenged by the sociolinguistic approach.

After structuralism/functionalism

The "baby-boom" of linguistic frameworks began in the 1950s, peaked in the 1970s and never slowed down. Most theories developed in the second half of the 20th century were based on the study of modern languages, namely:

- Hebrew and English (and later Romance and Greek) by the different subfields that derived from Chomsky's theories and are grouped under the label of "Universal Grammar".
- "Exotic" languages, both by generativists (e.g. Hale on native American languages, Huang on Chinese, etc.) who tried to expand their focus and by typologists who define structure and universals in a way different from the "UG(s)".

Contemporary historical linguistics

Contermporary historical linguistics, especially when applied to the Indo-European studies, has been rather impervious to the main "universal" frameworks, because very few frameworks apply "automatically" to text-languages.

Thus, most indo-europeanists work within a functionalist/ structuralist framework that changed very little over the last decades.

Part 2: Syntax and Formalism

Syntactic Frameworks

A syntactic framework is a partial framework or a part of a general framework dedicated to the assessment of syntax.

The most famous family of syntactic frameworks contains the ones that derive from the works Chomsky published starting from the 1950s. Think for instance of: general PSG, X-bar theory, LFG, HPSG, bare PSG, GBT, Standard Model, Minimalism, Cartography.

Other frameworks exist: Relational Grammar(s), Construction Grammar(s), Functional Grammars (e.g. Role and Reference Grammar), etc.

Criticism (1)

Syntactic frameworks can be - and have been - criticized in two ways:

- General criticism to the very idea of a framework (e.g. Givón 1984[1]; Haspelmath 2008).
- Specific criticism to a single model or to a group of related models (e.g. Van Valin 2005 & 2008 on UG).
- Criticism to a specific part of a model (e.g. Haspelmath 2004 on X').

Criticism (2)

Regarding general criticism to the very idea of a framework, most of the arguments proposed depend on some misrepresentations:

- A framework is not assumed to be unchangeable (therefore, it is not an "oppressive" environment);
- Every science uses frameworks; they map what Kuhn called "paradigms" over a specific metalanguage and are the way science proceeds;
- Universality is not a way to transform every language into English; universality is a working hypothesis based on the study of human brain and on the fact that languages are likely to share *a core* of common operations and structures (e.g.: *every child and no horse can learn any human language between the age of 0 and 4*; or again: *every predicate in any language can be negated"* are universal traits);
- In my opinion, the attacks to "configurationality" aimed at denying syntactic universals depend on a **strong definition of configurationality** and/or are based on a misconception of the very definition of "phrase".

Configurationality and Phrases

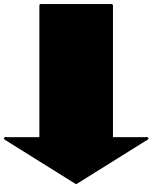
X'-languages vs. W-languages (Hale apud Chomsky 1981). Distinction **now mostly abandoned**. "Configuration" regards the hierarchic structure of phrasal composition, it is not necessarily *directly* reflected by the linear one.

Also note: if configurationality exists, it regards the hierarchic form of the logic structure of natural language. Thus, speaking of "onset" or "development" of configurationality is a contradiction in terms.

More on Configurationality and Phrases

Weak definition of Phrase

In my opinion, a "phrase" is *not* necessarily defined as a fixed position in clause architecture. In fact, **a phrase is any set of words with cardinality $|w^+|$ that is treated as an atomic object by at least one syntactic operation within a language** (= any set of words that is a phrase according to at least one valid constituency test).



Weak definition of Configurationality

"Configurationality" means that languages are structure-dependent. It should not necessarily imply fixed phrase positions within a universal clausal structure.

More on Configurationality and Phrases

Power of these weak definitions:

In order to challenge this weak definition of configurationality (= dependency from structure), one should prove *not just* that in Tagalog the order of constituents requires pre-topical predication, but that there exists a given language in which the positions of structurally related elements are entirely independent from each other.

This is intuitively hardly conceivable.

Part 3: Can/Should We Use Syntactic Formalism

Syntax exists

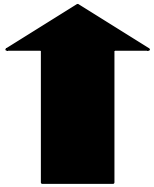
Configurationality, even a weakly defined one, implies that syntax exists. It is at least apparent as a set of consistent constraints that regulate the epiphenomenic (c-structure, s-structure, call it whatever you like) structure of the sentences in a given language.

Also, a model of syntax must be powerful enough to explain arbitrary recursive hierarchies (as per Chomsky, 1957).

Therefore, assuming that information-flow alone is sufficient to describe the structure of a language would be simply and utterly mistaken.

Caveat hostis

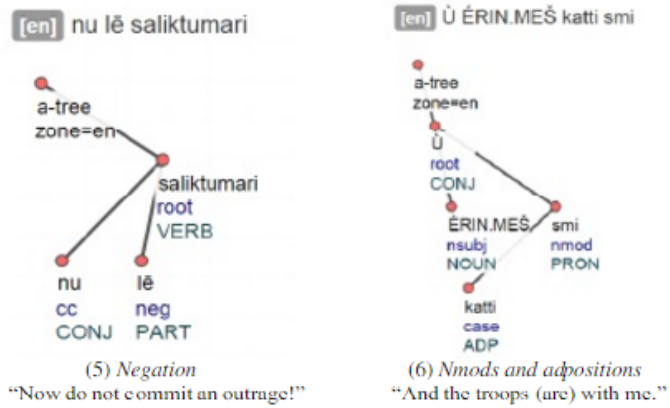
Therefore, assuming that information-flow alone is sufficient to describe the structure of a language would be simply and utterly mistaken.



This does not mean that informational markedness is not a part of syntax. It means that the kind of constraint that prevents locality violantions and the kind of constraint that prevents post-topical predication in a given language depend on a different level of organization, unless one can prove that they are the same.

Syntax exists → syntax must be represented

This also implies that formal notations that bypass the partial hierarchic order of syntactic posets (e.g. heads governing dependents) may be syntactically inconsistent. Consider for instance this way of annotating conjunctions/connectors.



(5) *Negation*
"Now do not commit an outrage!"

(6) *Nmods and adpositions*
"And the troops (are) with me."

This is a proposed UD-markup for Hittite (Inglese 2015). However, in (5) the connector *nu* (marked as conjunction but translated as an expletive) is dominated by the predicate, while in (6) the akkadographic conjunction *Ù* governs the whole copular clause.

An Example of Syntactic Constraint

How do syntax and information flow interact?

Consider the set of possible [Dem + Det + N] combinations in Ancient Greek*:

τοῦτο τὸ βιβλίον [syntactically possible (anaphoric) – focus on **demonstrative**]

τὸ βιβλίον τοῦτο [syntactically possible (deictic) – focus on "**book**"]

*τὸ τοῦτο βιβλίον [syntactically impossible]

---

*On the different explanations of this and similar phenomena cf. Bakker, 2009, 261ff.; Alexiadou et al., 2007, 127ff.

Back to formalism: what now?

We hopefully agreed that syntax exists, that it must be represented, and thus that some kind of formalism is required.

But what kind of formalism can we - and should we - use to represent the syntactic structure of Indo-European text-languages?

Top-down

vs.

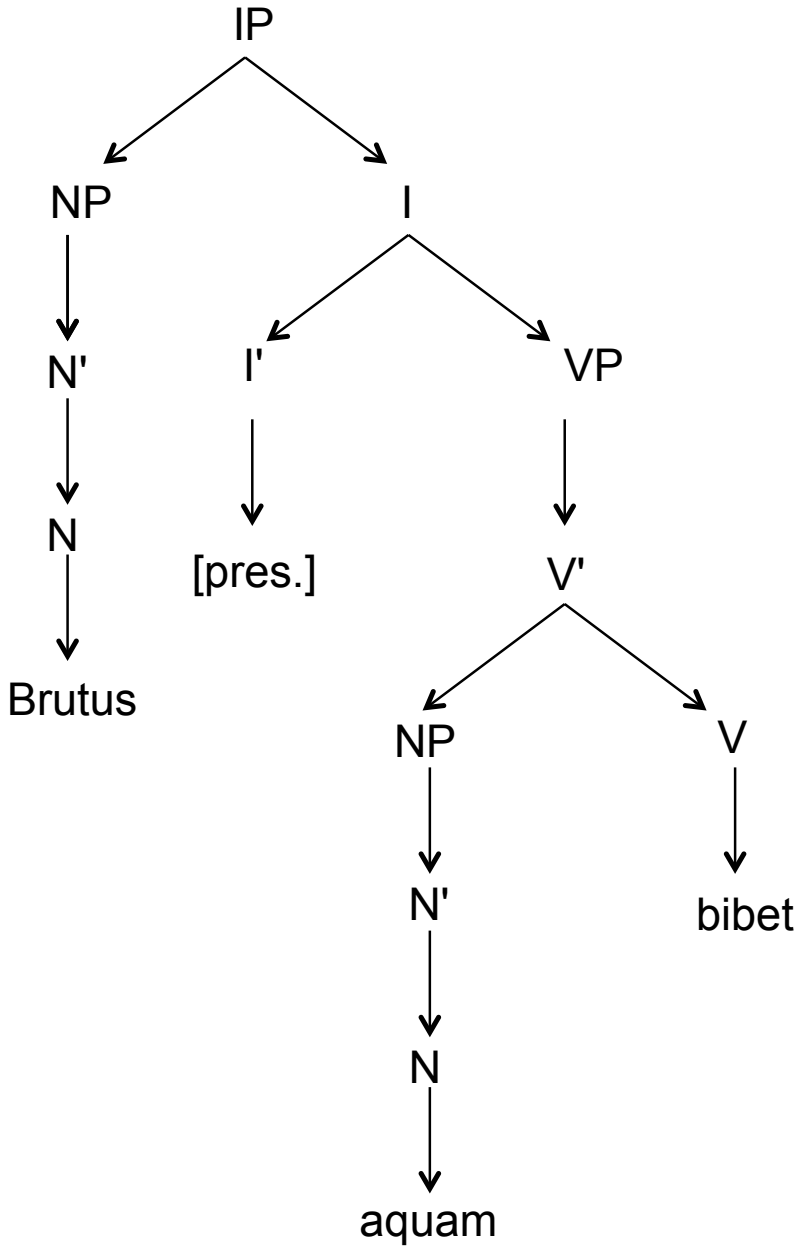
Bottom-up

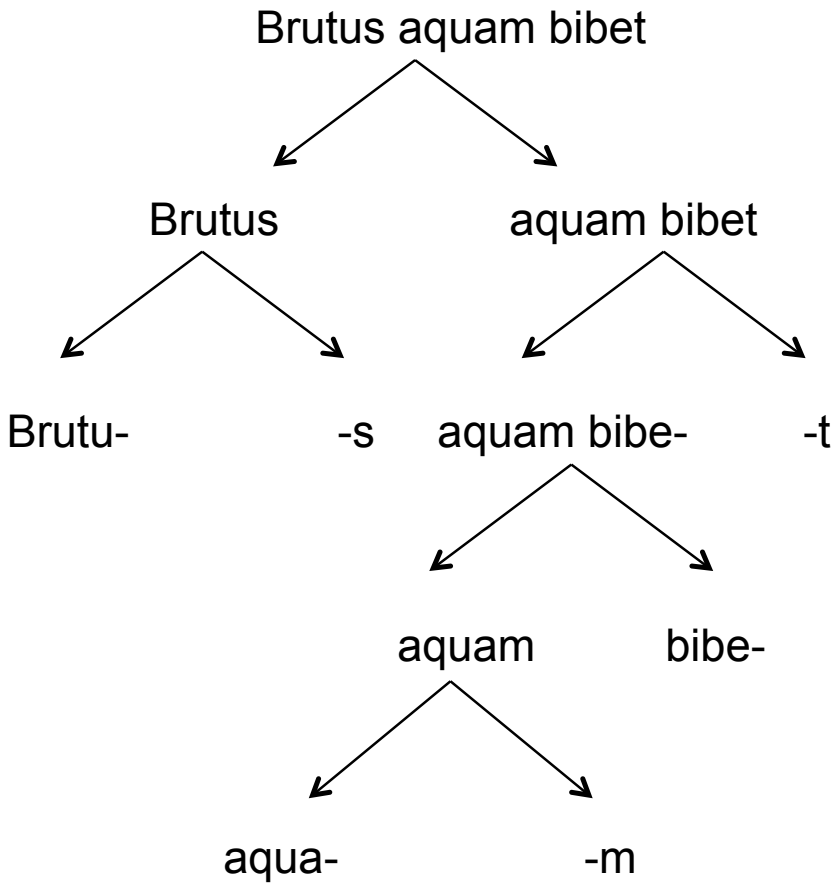
Most UG-related methodologies are presented as models that *admit* a top-down assessment (e.g., the generative grammars; note that LFG, for instance, is less rigid): I can start from a pre-linguistic "mental" phase and define the steps that bring to the linearization (and phonation) of the linguistic act ("**interfaces**"). This is especially true for the transformational ones.

Other methodologies can be presented as *strictly* bottom-up, they start from linguistic data (spoken or written corpora or examples) and must build a theory basing on them.

Since I cannot ask a Luwian whether the phrase *amanza za tarusa* would be less grammatical than *za amanza tarusa* (this statue of mine), it is pretty evidente that we must rely on the data we have, and look for a methodology of the second type.

Look at these trees





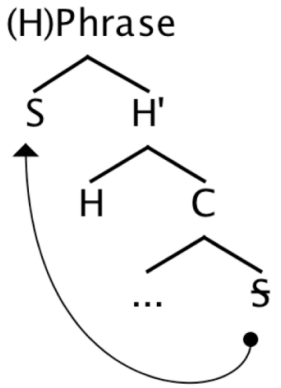
The two trees, and beyond

The one on the right starts from an actual sentence and deconstructs the syntactic steps of its formation mapping it over morphosyntactic indicators. The one on the left is a generative tree that works within the model of X' (and can be subsumed by quite a few later versions of the "generative" frameworks).

My point is that both trees can represent an ancient PIE language – but the first one contains theoretical hypotheses that are not verifiable with the study of a text-language.

The case of antisymmetry

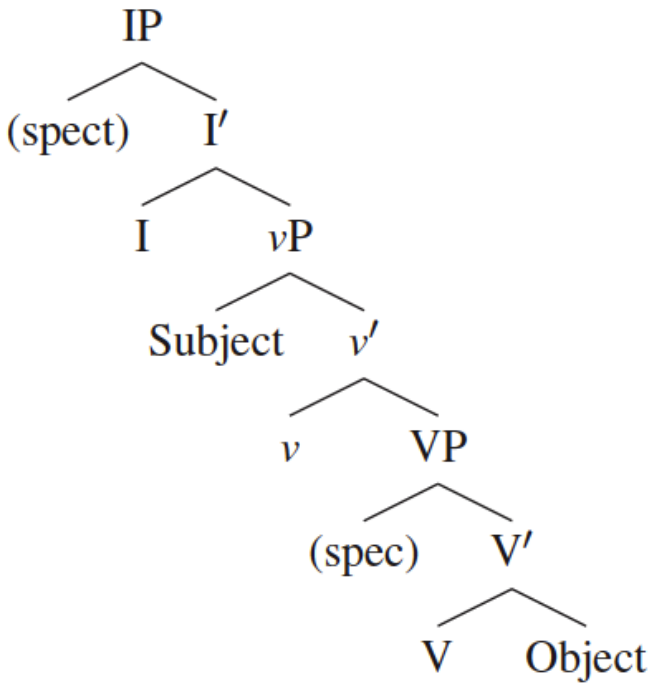
Antisymmetry is a good example of assumption that cannot be proven by the study of the syntax of a text-language, and in most cases integrating it can be difficult.



The antisymmetric theory was formulated by Kayne (1994). It predicts that the generative X'-trees of sentences in any language follow unary or binary branching and that the order is: specifier-head-complement. This allows for a simpler definition of locality constraints and movement rules; also, it certainly makes X' compatible to a left-regular formal language.

A famous consequence of antisymmetry

Antisymmetric theory dictates that, within the maximal projection of an inflected verbal phrase (which is usually labelled IP), the universal order of core constituents should be SVO:



Evidently, this consideration (which holds true even after Moro's proposal of a *Dynamic Antisymmetry*) collides with the typological data from the languages of the world: 45% of them – which means the realtive majority – has a SOV word order.

Thus, we must ask ourselves a question. Should we reconstruct a Deep-Structure SVO order when looking for an Anatolian, or Latin, or Sanskrit syntactic formalism?

So what?

Antisymmetric theory may be correct or may be wrong. But certainly it is not in the analysis of text-languages that we can find answers on generative models of language production.

As in the case of the two trees shown above (→ *look at these trees*), the problem is that every assessment of text-languages will always be bottom-up, and can only mark, parse and analyse what the available data represent.

All in all, representing movements for which we have no evidence is definitely irrelevant in historical syntax.

The dependency/constituency debate

Regarding the problem of *annotation* proper, dependency grammars (à la Tesnière) and constituency grammars (or PSG, à la Chomsky) have been competing in the last decades. However, provided that

(1) dependencies actually respect predicate/argument structure

(2) constituencies contain a unique head for each phrase

the two representations have been shown to be weakly equivalent and, for the purpose of a theory-free annotation, they both satisfy the requirements of a thorough syntactic annotation.

Summing up

- All languages are structure dependent (weakly configurational) and their structure does not depend on information-flow alone.
- Syntactic representation is thus a part of language description and cannot be neglected.
- Dependency representation is suitable as long as it respects phrase structure.
- Several aspects of generative syntax are not applicable to the bottom-up approach to formal representation of text-languages.
- In general, **syntactic formalism for text-languages should highlight the syntactic relationships and operations involving the phrase structure of the clause.** Under this perspective, both a correct dependency annotation and an unbiased phrase structure annotation can be employed, with a negelctable loss of information.

Basic References

- Alexiadou, A. et al., 2007, *Noun Phrase in the Generative Perspective*, Berlin/New York.
- Bakker, St., 2009, *The Noun Phrase in Ancient Greek: A Functional Analysis of the Order and Articulation of NP Constituents in Herodotus.* Leiden: Brill.
- Chomsky, N., 1957, *Syntactic Structures*, The Hague.
- Chomsky, N., 1959, On certain formal properties of grammars, in *Information and Control* 2, pp. 137-167.
- Chomsky, N., 1981, *Lectures on Government and Binding: The Pisa Lectures*, Foris Pub. Hollande.
- Givón, T., 1984, *Syntax, an Introduction*, Amsterdam (1st edition).
- Graffi, G., 2006, "20th Century Linguistics: Overview of Trends", in Keith Brown (editor in chief), "Encyclopedia of Language and Linguistics", 2nd ed., Oxford, Elsevier, vol. 13, pp. 181-195.
- Graffi, G., 2013: "European Linguistics since Saussure", in The Oxford Handbook of the History of Linguistics , ed. by Keith Allan, Oxford, Oxford University Press, pp. 469-484.
- Haspelmath, M., 2004, "Does linguistic explanation presuppose linguistic description?" Studies inLanguage 28, 3: 554-579 .
- Haspelmath, M., 2008, "Framework-free grammatical theory." In: Heine, Bernd & Narrog, Heiko (eds.) The Oxford handbook of grammatical analysis. Oxford: Oxford University Press, 341-365.
- Inglese, G., 2015, "Towards a Hittite Treebank. Basic Challenges and Methodological Remarks", in Proceedings of the workshop *Corpus-based Research in the Humanities*, Warsaw 2016, pp. 59-68.
- Kayne, R.S., 1994, *The Antisymmetry of Syntax*, Cambridge MA.
- Moro, A., 2000, *Dynamic Antisymmetry*, Cambridge MA.
- Schneider, G. 1998, *A Linguistic Comparison of Constituency, Dependency and Link Grammar*, Thesis Universität Zurich, 1998.
- Tesnière, L. 1959, *Éléments de syntaxe structurale*, Paris.
- Universal Dependencies: https://universaldependencies.github.io/docs/
- Van Valin, R.D. Jr., 2005, Exploring the Syntax-Semantic Interface. Cambridge: CambridgeUniversity Press.
- Van Valin, R.D. Jr., 2008, Role and Reference Grammar. In: Bernd Heine and Heiko Narrog(eds.), The Oxford Handbook of Linguistic Analysis. Oxford: Oxford University Press.