Borut Pretnar

On algebraic aspects of syntactic relations

Abstract:.A formal approach to sentence and phrase decomposition is proposed. It is based mainly on the observed analogy between mathematical operations performed on algebraic objects on one side and joining of syntactic constituents to larger composite constituents on the other side. The working elements of algebraic operations (operators and operands) are mapped on syntactic constituents. The proposed approach is partly motivated by two existing (but disparate) ideas: first to treat syntax as an algebraic object and second by the attribution of different roles (asymmetry) to joined constituents. In the proposed approach both aspects are unified and coherently integrated. A graphical representation of sentence decomposition in matrix form is proposed and demonstrated as a suitable representational tool.

Keywords: syntactic-algebraic analogy, operator-operand relation, asymmetry of constituents, matrix representation.

1. Introductory remarks

The building blocks of sentence (constituents) are traditionally studied by syntax, a branch of linguistics. In the past decades, several different approaches to sentence decomposition have been proposed, among them notably the phrase structure grammar. Table 1 shows some relevant approaches in approximate chronological order. The approaches based on asymmetry of joined constituents (i. e. with different roles attributed to involved constituents) and the approaches including algebraic interpretations are highlighted and commented.

A brief terminological comment is a necessary prerequisite. As noted in Table 1, joined syntactic constituents have been approached and denominated from different theoretical viewpoints. Some traditional approaches as e. g. »*immediate constituents«,* »*phrases*« partitioned in »*heads*« and their »*complements*« and recently »*merger*« (Cook and Newson 2007: 250) do not take notice of the asymmetrical relation between the requiring and the requested constituent explained in the next section. However, the followers of the so called »*dependency grammar*« do indicate this kind of asymmetry using the notions of »governor« or »valency carrier« as opposed to »*dependent*«. Trask′s dictionary (Trask 1992: 195) mentions the notional pair »*operator*« and »*scope*«.
In the present article the terms »*operator«* and »*operand«,* based on the algebraic-linguistic analogy demonstrated in the next section, are consistently used.

Item No.	*Description and/or denomination of syntactic approach*	*In the role of operator*	*In the role of operand*	*Brief comment and source*
1	Immediate constituents	No asymmetry		(Bloomfield 1933; Lyons 1968: 210-212)
2	Categorial grammar	The asymmetry is adapted to match the constraints of the so called »*connectivity*«.		Introduces the important notion of *algebraic operation* for joining constituents. (Ajdukiewicz 1935; Bar-Hillel 1953; Lyons 1968; 227-231)
3	Phrase structure grammar (rewrite rules »*trees*«)	Partition of *phrases* into *heads* and their *complements.*		(Chomsky 1957; Lyons 1968; 215-227; Cook and Newson 2007; 28-32; Tallerman 2015: 114-150)
4	Advanced generative grammar	Asymmetry defined by *c-command*		(Dürscheid 2012: 125-152)
5	Dependency grammar	*Governor (Regens)* *Valency carrier*	*Dependent*	Compares the asymmetry of the joined constituents to chemical valence bonds. (Tesniere 1959; Dürscheid 2012: 106-124)
6	The indicative chapter title: »Syntax as a foundation of intelligence«	Verb in the predicative position as active operator	Subject, object, adverbials etc.in the passive role as operands	Mental syntactic operations represented by the pictorial »Calvin Vacuum-Lifter Package-Carrying System«. An account of the cognition-based operator-operand syntactic asymmetry. (Calvin 1998)
7	Devlin′s basic thesis (Devlin 2001; 2, 70) »The features of the brain that enable us to do mathematics are the very same features that enable us to use language«.	No asymmetry		It is assumed that the ability of »doing mathematics« is a byproduct of the ability to use and understand syntax - due to an assumed cognitive parallelism between them. By »doing mathematics« is meant abstract algebra with the notions of e. g. «relation«, »operation«, «group« etc. explained at great length in the popular book. For an explanation of algebraic terms see e. g. (Cheng 2023:107; Partee et al. 1993: 27-36, 39-51, 247-252, 255-274, 546-551)

Table 1: Selected syntactic approaches with comments, especially on algebraic interpretation and the asymmetry relation. The quoted versions of asymmetry differ from the asymmetry approach proposed in this article.

The distinction between the *requiring and requested constituents of a composite constituent* has been actually known and described since a long time in different contexts and by different terms, notably by (Tesniere 1959; Calvin 1998) and many followers of the ideas subsumed as *dependency grammar* (Dürscheid 2012; 106-124). Additionally, the initiative to interpret the joining of linguistic constituents algebraically is due to early works of (Ajdukiewicz 1935) and (Bar-Hillel 1953) and more recently to (Devlin 2001). Both approaches are used as starting points for the integration into a new coherent approach as described in the next section 2.

In section 3, the traditional graphical representations of sentence structure, i. e. the bracketed representation and »*trees*«, are discussed and compared to the proposed matrix (tabular) representation, arranged in rows and columns.
Section 4 concludes with some general comments on possible implications of the proposed approach.

2. Analogy between algebraic and linguistic operations

The comments in Table 1 summarize the characteristics of important traditional approaches in order to make them comparable to the approach proposed below.

Consider the structure of two very simple algebraic operations, i. e. addition and multiplication:

[1] Number(1) **+**(»plus«) Number(2) = Sum
[2] Number(1) **X**(»times«) Number(2) = Product

and compare them with the linguistic operation of joining constituents to a composite constituent:

[3] Constituent (1) and Constituent (2) to be joined to (Composite constituent)

Note that in the composite constituent, the original constituents may (as a rule) remain separated or become integrated into a composite word.

The algebraic operations [1] and [2] consist of two easily identified elements: numbers as *operands,* subject to an algebraic operation (addition or multiplication) and *operators »plus« and »times«,* symbolized by the signs »**+**« and »**X**«. Operators function as »*commands*«, telling how to act, i. e. to transform (to »map« in the mathematical jargon) operands to something else, here sum or product, respectively. In addition, operators can be semantically identified as the »*active*« and the operands as the »*passive*« constituents.

It is interesting to investigate what can possibly serve as operator and operand in the *linguistic* case [3]. At first glance it seems that both constituents (1) and (2) are operands and the »command« to join them is the actual operator. This pattern agrees superficially with the above algebraic cases [1] and [2].
However, a thorough reflection shows that there is a plausible alternative possibility: one of the constituents is operator and the other is operand. This idea resembles the approaches known as *»dependency grammar«* and *Calvin's ideas*, items No. 5 and 6

in Table1. The following discussion exposes the arguments supporting this alternative and the consequences thereof. For brevity, the arrow symbol »→« is used in the subsequent text. It symbolizes the asymmetric relation, pointing »from operator → to operand«.

First note the different properties of operators vs. operands in the algebraic case. In absence of operands, the bare operators are completely senseless and ineffective because their intrinsic function to produce a result (e. g. a sum or a product) can not be activated. This is not true for operands. Numbers can and do serve as stand-alone items, e.g. as page numbers, results of counting etc. Due to these different properties, operators may be identified as the *requiring* and operands as the *requested* constituent.

This characteristic difference can in many cases find its counterpart in syntactic structures: e. g. adjectives and prepositions are not functional as stand-alone items, they require nouns (or noun phrases) to activate their function. Nouns, in contrast, are functional as stand-alone items, e. g. as names or titles. Thus it is possible to use characteristic different properties to attribute the *operator→operand* roles to the constituents of a joined composite constituent. Constituents *requiring* another constituent take the *operator role* while *requested* constituents, usually capable to be a stand-alone item, take the *operand role*. This statement can serve as a powerful (although not universally applicable) test. Most syntactic cases are analyzable in the proposed asymmetric *operator→operand* way as demonstrated in Table 2. However, there remain ambiguous cases, some of them discussed below, where additional semantic criteria may help to resolve the questions about the role assignment. Note that the semantic relation »*operator increases the information about the operand*« is confirmed in some important cases: [Adjective →Noun], [Preposition→Noun] and [Predicate→Subject].

Operators	*Operands*	*Remarks*
Verb (V) as predicate	Noun as subject (argument)	The so called »null subject sentences« are rare exceptions
Transitive V	Noun (or multiple nouns) as object(s)	
Modal verbs, auxiliary verbs, »do«, some special verbs like want, begin	Verb as infinitive	Auxilliary »do« as used in questions and negations
Auxiliary »have«	Verb as past participle	
Phrasal and prepositional verbs	Preposition	
Special class of verbs	Noun or Adjective as predicate	e. g. to be, to become, to remain
Noun in attributive role	Noun	German compound nouns consisting of »Bestimmungswort« and »Grundwort« are paramount examples.
Determiner	Noun	
Adjective »modifier«	Noun	
Compound attributes	Noun	Compound attributes may be e. g. genitives indicating properties or possession of a noun or whole relative clauses referring to a noun.
Preposition	Noun	
Adverb (Adv) »modifier«	Verb and Adjective	
Negation (NOT)	Verb and Adjective (Noun limited only)	Limited to special cases of noun as e. g. nonresident; nonsmoker; nonsense
Complementizer »that«	Subordinate clauses	
Conjunction »and«	Several classes of words, phrases and clauses	Binary operands

Table 2*:* Frequent occurrences of *operator→operand relation*, taken mainly from English syntax. The table is not exhaustive and the examples may vary in different languages**.** Linguistic terms in widespread use (modifier, argument, attribute …) are added where appropriate. In many cases the abbreviations, e. g. N, V imply also the corespondent phrases NP, VP.

Constituents appended to verb (with the exception of subjects) may present ambiguous situations. (Dürscheid 2012: 106-124) summarizes the controversial opinions about the distinction between obligatory and optional constituents (complements and adjuncts) appended to verbs, published mainly in the German linguistic sources. Table 3 serves as a general guide here, an in-depth discussion exceeds the scope of this article.

Operators	*Operands*
Verb (predicate)	Obligatory appended constituents (rendering sentence ungrammatical, if omitted)
Optional appended constituents (not producing ungrammatical sentences, if omitted)	Verb (predicate)

Table 3: Different roles of obligatory and optional constituents (objects, prepositions, adverbials) appended to verb.

Composite constituents containing numerals may pose some further questions, e. g. in [4]:

[4] *Five men*

»Five« is formally a number and as such can exist as a stand-alone item, and so can the noun »men«, too. However, the numeral in this context is, semantically, a counting device requiring counted items as operands and is therefore an (attributive) operator.

There may exist additional ambiguous cases to be revealed in critical assessment of the proposed approach.

Note that sentence structure based on *operator→operand relation* is not compatible with the *phrase grammar* structure because *operators* and *operands* in general do not coincide with *heads* and their *complements*, respectively. As a consequence, the sentence decomposition (parsing) differs from traditional *»trees*« as described in the next section.

3. Alternative representations of sentence decomposition

In order to graphically display syntactic structures, decomposition of sentences makes use of several tools, e. g. rewriting rules, phrase structure tree diagrams (»trees«), hierarchically arranged brackets, parallel horizontal lines extended in phrase projection ranges (Morenberg 2002: 279-336) and specially devised algebraic symbols e.g. (Bar-Hillel 1953). »Trees« play a distinguished role in syntax, they serve also as the base for the definition of important syntactic notions, e. g. *c-command* (Dürscheid 2012, 132).

It may not be widely known that the notion of *relationship* can be treated as algebraic object (Cheng 2023: 17-18, 52-66, 82-94; Partee et al. 1993: 27-36). Relations are a key concept in sentence decomposition. While the traditional »trees« and brackets are based on the relation »*consists of*« (or, in the reverse direction, »*is part of*«) between the consecutive levels, the asymmetric decomposition introduced in section 2 is based on the relation »*operates on* → « (or, in the reverse direction, »← *is operated by*«). This difference has a profound influence on the pattern of decomposition as demonstrated on a simple sentence example [5]. It is first decomposed in the bracketed versions to demonstrate the difference between the traditional decomposition [6] and the decomposition based on *operator→operand relation* [7]:

[5] *Words convey the meaning.*

Traditional decomposition according to phrase structure grammar:

[6] {($_{N}$ Words) [$_{VP}$ ($_{V}$ convey) ($_{NP}$ ($_{D}$ the) ($_{N}$ meaning))]}

Decomposition according to *operator→operand relation* displays additional information:
[7] {($_{N}$ Words) ←[$_{VP}$ ($_{V}$ convey) →($_{NP}$ ($_{D}$ the) →($_{N}$ meaning))]}

For further examples of decomposition the matrix form of representation (Tables 4 and 5) is used. It is flexible, i. e. unlike the »*trees*« and bracketed decomposition it can disregard the linear (consecutive) word order in sentence. This is possible because the sequence of rows or columns in the matrix can be changed without loss of information, Such flexibility is welcome if the linear word order is irrelevant or intentionally discarded (Cook and Newson 2007: 268-269; Jackendoff 2011: 599). Additionally, It is particularly well adapted for the display of the operator→operand relation:

Operators → *Operands ↓*	VP (convey the meaning)	V convey	D the
N Words	X		
NP (the meaning)		X	
N meaning			X

Table 4: Sentence [5] decomposed in matrix form according to *operator-operand relation.*

For the sake of completeness and comparison it is demonstrated that the traditional »*trees*« can be converted to matrix representation, too, if appropriate for whatever reason:

Constituent → Consists of: ↓	Sentence	VP (convey the meaning)	NP (the meaning)
N Words	X		
VP (convey the meaning)	X		
V convey		X	
NP (the meaning)		X	
D the			X
N meaning			X

Table 5: The traditional »*tree*« of the sentence [5] converted to matrix form.

The comparison of tables 4 and 5 emphasizes again the difference between the two approaches. Note that the initial constituent in Table 5*,* »*sentence*«, is neither operator nor operand and does not appear in Table 4. The binary nature of the classical decomposition is particularly well visible in Table 5.

There is a strong additional reason for the matrix representation of the *operator→operand relation*. In some cases, two (or more) arrows point from the same operator to several simultaneous (not consecutive !) operands, rendering the bracketed representation inappropriate. Such cases include ditransitive verbs with double obligatory objects, e. g. *give* (Tallerman 2015: 41-43) *.* An additional case occurs in languages using serial verbs (Tallerman 2015: 102-104) where simultaneous arrows point from multiple predicates towards a single subject. Such occurences are obviously better handled by matrix representation.

4. Concluding remarks

Syntactic operation performed by an operator on an operand may have a counterpart in a cognitive physiological event in the brain (Calvin 1998; Devlin 2001). An experimental test of this hypothesis may present a challenge for cognitive psychologists.

The *operator→operand relation* described above allows some speculations about language evolution. Since the availability of operands is a necessary condition for operators to exert their function, operands must have existed before operators (or, in the improbable case, arise simultaneously with them). Thus e. g. nouns may have been a much earlier language feature then verbs and our early ancestors may have used noun phrases as a (semantically limited) means to communicate. (Aitchison 1996: 111) confirms that »nouns alone might be useful as a communication device« and that »…verbs came before nouns is unlikely«. In (Jackendoff 2002, 257-259) it is demonstrated how close the information contents of similar noun phrases and verb phrases can actually be.

Human language is a flexible means well adapted to name and combine real and imagined objects, concepts, situations, and processes and capture their essence ("*map*" them) into words and sentences. In mathematical terms, this ability could be denominated as "*mapping ability*" or, in Jackendoff's words (Jackendoff 2002: 257), "*expressive power*". It may be due mainly to the flexible role of the verb (as predicate), capable of multiple simultaneous roles as operator and operand (e. g. in relation to subject, obligatory complements and optional adjuncts) and to nouns, stored in set-like structures and semantic networks in the human memory (Solso et al. 2008: 254-259; Widdows 2004).

The current development in mathematics, particularly in algebra, may exert further influence on linguistics. The traditional well known set theory (even in its "naïve" form), often coupled with abstract and linear algebra, has served well for a long time and still does so e. g. (Widdows 2004). However, the classic situation of elements, sitting passively inside a set with the possibility to become operands of a prespecified operation (or a few of them) was changed by the advent of *category theory* in the forties of the past century. It brought a fresh perspective and was welcomed as a promising development by some mathematically minded linguists (Partee et al. 1993: 252). Here we have to deal with "*objects*", some of them in pairs equipped with inherent directional (asymmetric) interactions (called "*morphisms*" or just plain "*arrows*"). This situation remarkably resembles linguistic syntactic situations, e. g. the structure of decomposition in the bracketed example [7]. However, a caveat is appropriate: an exact mapping of sentences on a genuine algebraic category is, in general, not possible without violating (at least some) axioms of algebraic categories, quoted e. g. in (Cheng 2023: 102).

Last not least, dealing with mathematical aspects of language rises some tough and possibly (presently or forever) unanswerable questions, concerning linguists as well as mathematicians, particularly information theorists:

- Is it theoretically possible to define a (scalar or multidimensional) parameter, able to measure the above described "*mapping ability*" and to express e. g. the distance between a pidgin and a fully developed language in quantitative terms?
- Is it theoretically possible to devise alternative constituents of a "language" (possibly replacing verbs and nouns), but with equal or better "*mapping ability*"? In other words, can linguistics be turned from a descriptive to an engineering branch of science?

Abbreviations: A Adjective, Adv Adverb, D Determiner, N Noun, P Preposition, V Verb; added P (as e. g. NP, VP etc.) denotes the corresponding phrase.

References

Aitchison, Jean. 2000. *The Seeds od Speech Language Origin and Evolution.* Cambridge: Cambridge University Press, Canto edn.

Ajdukiewicz, Kazimierz. 1935. “Die syntaktische Konnexität” *(Syntactic Connectivity). Studia Philosophica* 1. 1-28

Bar-Hillel, Yehoshua. 1953. “A quasi arithmetical notation for syntactic description.” *Language* 29. 47-58

Bloomfield, Leonard. 1933. *Language.* New York: Holt, Rinehart & Winston.

Calvin, William H. 1998. *How Brains Think   Evolving Intelligence, Then and Now.* London: Phoenix, a division of Orion Books Ltd.

Cheng, Eugenia. 2023. *The Joy of Abstraction An Exploration of Math, Category Theory, and Life.*
Cambridge: Cambridge University Press.

Chomsky, Noam. 1957. *Syntactic Structures.* The Hague: Mouton.

Cook, Vivian J., Mark Newson. 2007. *Chomsky's Universal Grammar An Introduction.*
Oxford: Blackwell Publishing 3rd edn.

Devlin, Keith. 2001. *The Math Gene   How mathematical thinking evolved and why numbers are like gossip.*
New York: Basic Books.

Dürscheid, Christa. 2012. *Syntax   Grundlagen und Theorien (Syntax Foundations and Theories)*.
Göttingen: Vandenhoeck & Ruprecht 6th edn.

Jackendoff, Ray. 2003. *Foundations of Language   Brain, Meaning, Grammar, Evolution.*
Oxford: Oxford University Press.

Jackendoff, Ray. 2011. “What is the human language faculty? Two views.” *Language* 87. 3. 586-624

Lyons, John. 1968. *Introduction to Theoretical Linguistics.*
Cambridge: Cambridge University Press.

Morenberg, Max. 2002. *Doing Grammar*.
Oxford: Oxford University Press 3rd edn.

Partee, Barbara, Alice ter Meulen and Robert E. Wall. 1993. *Mathematical Methods*

in Linguistics.
Dordrecht: Studies in Linguistics and Philosophy 30, Kluwer Academic Publishers.

Solso, Robert L., Otto H. MacLin, M. Kimberley MacLin. 2008. *Cognitive Psychology.*
Boston: Pearson Education Inc. 8th edn.

Tallerman, Maggie. 2015. *Understanding Syntax*.
London & New York: Routledge 4th edn.

Tesnière, Lucien. 1959. *Eléments de syntaxe structurale.*
Paris: Klinksieck.

Trask, Robert L. 1992. *A Dictionary of Grammatical Terms in Linguistics.*
London & New York: Routledge.

Widdows, Dominic. 2004. *Geometry and Meaning.*
Stanford: Center for the Study of Language and Information.

About the author:

Borut Pretnar

ORCID: 0000-0001-6004-8522

Independent (no formal affiliation), long-standing member of Linguistic Circle at Philosophical Faculty of University of Ljubljana, Slovenia, with main interest in mathematical linguistics.
Linguistic Circle is an academic institution with decades long tradition and with »open access« membership policy. Until his death in April 2024 it was run under the professional auspices of academician prof, Janez Oresnik.

Authors educational background is metallurgical engineering and physical chemistry.