Algorithms, Representations, and ATR Ratios

Gereon Müller (Universität Leipzig) June 8, 2015

1. Introduction

In this short meta-linguistic paper, I pursue two related goals. First, I suggest that there has been a move away from more algorithmically oriented work on syntactic theory to more representationally oriented work. And second, I argue that this claim can independently be justified by invoking a new technique of measuring algorithmicity in syntactic analyses, viz., what I call ATR Ratios in example environments of journal articles.[1]

2. Algorithms vs. Representations

Suppose that syntactic theory is conceived of broadly, as a formal approach that systematically predicts the wellformedness or illformedness of linguistic expressions like phrases and sentences in a language, and is prepared to envisage abstract concepts in doing so. The field of linguistic study thus defined is characterized by a dichotomy of two concepts, viz., *algorithms* on the one hand and *representations* on the other, in the sense that these two concepts motivate and characterize two complementary perspectives on the object of study (structured linguistic expressions). The two perspectives are specified in (1).

(1) a. *Representational perspective*:
What does the structure of a given linguistic expression look like?
b. *Algorithmic perspective*:
How is a linguistic expression generated (or licensed)?

To give an arbitrarily chosen, extremely simple example: From a representational perspective, it can be asked what the syntactic structure is that underlies a string like *What did she see?*. Questions that need to be addressed from this perspective include whether there is an abstract item (e.g., a trace, a copy) in a position next to *see*, what position *did* shows up in (and whether it has been moved to that position or not), what type of position the wh-phrase *what* occupies, how many (if any) non-overt functional categories there are in this sentence, in what order they occur, and so on and so forth. In contrast, from an algorithmic perspective, it needs to be clarified how the string (together with its structure) emerges as a legitimate grammatical object in the first place; in doing so, questions need to be addressed for a sentence like *What did she see?* that involve the trigger for obligatory wh-movement, the necessity of carrying out *do*-support, case assignment to subject and object DPs, agreement of the subject with the auxiliary, subject-auxiliary inversion, etc. More generally, from an algorithmic perspective, the main focus is on what the *building blocks* of grammar (i.e., the rules, constraints, principles, elementary operations, etc.; cf. Stechow & Sternefeld (1988)) look like that generate (or license) these structures, and how they interact.

In the history of modern syntactic theory since its inceptions (say, Chomsky (1957)) , one can identify repeated changes in emphasis on algorithmic vs. representational aspects of syntactic theory. This to-and-fro has undoubtedly been very beneficial to the field: After a phase of mainly representational research that has led to better, typically more articulated structural analyses of linguistic expressions in syntax, a subsequent phase of research that is predominantly algorithmic

[1] The material that follows was part of a statement in the session on *Central research areas in generative syntax*, delivered at the Athens Conference on *Generative Syntax in the Twenty-First Century: The Road Ahead* on May 28, 2015.

can make use of the new structures and thereby improve on the core concepts employed in the theory; this then triggers a new phase of representational research; and so on.

Against this background, it can be noted that there have been at least four major developments since the early to mid 90s of the last century which would each seem to strongly support a focus on representational aspects of syntactic theory, viz. those in (2).

(2) *Models supporting a focus on representational aspects*:

a. Antisymmetry (Kayne (1994))
b. Cartography (Rizzi (1997), Cinque (1999), Belletti (2004))
c. Nanosyntax (Starke (2009))
d. Transparent Logical Form (Stechow (1993), Heim & Kratzer (1998))

What is more, it seems to me that the focus on representational aspects induced by these models has been accompanied by a decrease in work on algorithmic aspects in all four cases.

Thus, whereas the core of Kayne's antisymmetry approach (i.e., the Linear Correspondence Axiom; LCA) as such firmly belongs to the algorithmic realm of syntactic theory and has triggered some further work from an algorithmic perspective (e.g., Chomsky (1995), Uriagereka (1999)), it seems fair to conclude that the vast majority of antisymmetry-based work over the last decades has centered around suggesting and arguing for LCA-compatible structures (including various kinds of movement traces required by these structures), while at the same time neglecting, at least as a tendency, the question of how these structures can come into being in the first place (some exceptions like Koopman & Szabolcsi (2000) notwithstanding).

Similarly, the cartography enterprise has led to the postulation of many fine-grained structural distinctions. It has thereby hugely contributed to the understanding of syntactic representations; however, abstracting away from a few exceptions (see, e.g., Rizzi (2004))), the conclusion would seem to be uncontroversial that these analyses arguing for cartographic structures tend not to spend as much time showing how these structures are generated (licensed) in the first place, and what repercussions these structures have for well-established concepts of the algorithmic part of the grammar.

Arguably, similar conclusions hold for much of the work in nanosyntax (again, with exceptions; see, e.g., Caha (2009), which also has a focus on algorithmic aspects).

Finally, it seems to me that an almost exclusive emphasis on representational aspects also permeates a lot of the work that has been carried out on the syntax/semantics interface on the basis of a transparent logical form approach in recent years. In fact, whereas at some point there was quite a bit of discussion of the question of whether building blocks proposed for syntax (like the Empty Category Principle (ECP; Chomsky (1981)), the Path Containment Condition (Pesetsky (1982)), the Subjacency Condition (Chomsky (1977)), or Transderivational Economy (Chomsky (1991))) would also constrain the operations required to transform syntactic representations into transparent logical form representations that can directly form inputs to semantic interpretation (see, e.g., May (1985), Stechow (1996), and Fox (2000)), such work seems to have all but stopped. The predominant approach in this area nowadays rather seems to be to suggest articulated syntactic representations that best correspond to the semantic analysis adopted (given general guidelines like compositionality), even at the cost of introducing otherwise (i.e., purely syntactically) unmotivated categories, structures, and operations.

Not to be misunderstood, I firmly believe that all this research on representations is necessary, and has led to a much better understanding of the structure of linguistic expressions. Still, from the point of view of grammatical theory (more generally, of research on what constitutes the faculty of language), the core question is ultimately not so much what the structures of individual sentences

look like, but rather, what the building blocks (in the sense specified above) are that generate (or license) these structures, and how these building blocks interact. Thus, if these latter questions are not being pursued in sufficient depth, there is a danger that syntactic analysis will remain on a descriptive level, much like the approach to syntax in terms of fine-grained topological fields that plays such a big role in traditional research on Germanic languages (see, e.g., Höhle (1986)).

Assuming this assessment of the present situation in syntactic theory to be on the right track, it might be taken as an incentive to put some more emphasis on algorithmic research again that will then accompany, as an equal partner, the representational research that currently dominates the field.[2] But *is* this assessment really basically correct? Is there any independent, objectively verifiable method to substantiate the claim that by focussing on the *representational perspective* on syntax, the *algorithmic perspective* has been somewhat neglected?

As far as I can see, no such method has been proposed so far. With the present paper, I would like to remedy this situation.

3. Measuring Algorithmicity by ATR Ratios

I would like to suggest a new, extremely simple means of measuring algorithmicity in syntactic analyses. The background assumption is that when an algorithmic perspective is important in a syntactic analysis published in a scientific paper (or in a thesis, or in a book), a sufficiently precise formulation (or definition) of the building blocks that are involved will very often be unavoidable (given general requirements in the field of syntactic theory concerning coherence and verifiability); and I would like to contend that the most natural way to do this (given the established principles of writing papers and monographs in this field) is to separate the formal building block characterization from the rest of the text by putting it into an example environment and numbering it, just like object language examples. On this basis, the idea is to look at the ratio of *algorithmic material* (building blocks of grammar: principles, rules, constraints, operations, extragrammatical factors, etc.) and *representational material* (example sentences, structured expressions, trees, etc.) in numbering sequences (i.e., complete sequences of example environments) of journal publications. I call the resulting value the ATR Ratio (Algorithmic material To Representational material Ratio in numbering sequences).

The procedure as such looks as follows.

(3) *ATR Ratio*:

a. Determine the numbered items (example environments) of a journal article.
b. Register those which are not actual object language examples, structures, trees, etc., but concerned with genuine building blocks of grammar.
c. Determine the ratio of these instances of numbered building blocks to all numbered items in a journal article.
d. Do this for each article (in syntactic theory) in a journal.
e. Determine the overall ratio of algorithmic material to non-algorithmic material in a

[2] Note that the issue of an algorithmic vs. representational approach must be kept apart from another question, that of whether the the algorithmic part of syntax is *derivational*, with a sequential application of building blocks, or *declarative* ("*representational*"), such that all building blocks apply simultaneously – or, for that matter, whether the algorithmic part of syntax is characterized a mixed approach. Arguments for a derivational organization of syntax typically come from opaque interactions (counter-bleeding and counter-feeding), and it seems to me that the interactions of, say, reflexivization and movement, or movement and deletion, or movement and case assignment, do indeed ultimately favour a derivational approach (see Müller (2014)). However, this issue is entirely orthogonal to the question of an algorithmic vs. representational perspective. (This is reflected in the decision to avoid a choice between *generation* (= derivational approach) and *licensing* (= declarative/representational approach) in (1).)

journal in a year.

A tacit presupposition here is that material in numbered items that is not algorithmic in nature is almost exclusively representational. At least for the case study reported below, this is undoubtedly the case.[3]

4. Case Study

As a case study, I looked at the numbering sequences in five volumes of *Linguistic Inquiry*: 1989, 1993, 1994, 2013, 2014. This choice is mainly due to the fact that *Linguistic Inquiry* is arguably still the most influential of all linguistic journals with a focus on syntactic theory; in line with this view, it also emerges as the best-ranked journal in this area from a scientometric perspective (e.g., it outranks *Natural Language and Linguistic Theory* and *Syntax* in the Scimago Journal Ranking for 2013, with *Language*, *Lingua*, *Linguistic Review*, *Journal of Linguistics*, *Theoretical Linguistics* and *Linguistics* (in that order) placed much further down the list of journals that regularly publish work in syntactic theory. Also, *Linguistic Inquiry* would seem to have a reputation for focussing on *theoretical* rather than, say, *empirical* or *descriptive* work, which makes it particularly suited to the task at hand: If anywhere, this is where one would expect to find algorithmic material in abundance.

The five volumes were chosen based on a number of different criteria: The 2013 and 2014 volumes represent the situation as it presents itself now; the 1993 and 1994 volumes predate two theoretical innovations which have led to an increased interest in the nature of syntactic representations (viz., antisymmetry and cartography); and the 1989 volume predates the first versions of the minimalist program. Of these five volumes of *Linguistic Inquiry*, only those articles were considered which have a main focus on syntactic theory (i.e., purely phonological, morphological, or semantic studies were excluded). In addition, material that does not involve Arabic numerals was not considered. This includes material accompanied by Roman numbering in footnotes, and material separated from the main text in some other way, e.g., by simple indentation – although it should be noted that this latter kind of strategy seems to be virtually absent from *Linguistic Inquiry*, especially with building blocks.

The results of the case study are given in (4).

(4) *ATR Ratio in five volumes of Linguistic Inquiry, 1989-2014*

Vol.	algorithmic material	all numbered items	ATR Ratio
2014	54	1194	**4,5**%
2013	36	1040	**3,5**%
1994	121	1105	**11,0**%
1993	142	1353	**10,5**%
1989	137	968	**14,2**%

As illustrated in (4), the overall number of example environments does not vary drastically in the five volumes under consideration; the number of example environments addressing building blocks, however, differs substantially, and so does the ATR ratio emerging here. Thus, there has been a clear decline of algorithmicity over the years, with the ATR ratio moving from 14,2% in 1989 via 10,5–11,0% in 1993, 1994 to a meagre 3,5–4,5% in 2013, 2014.

The number of articles that could be considered (i.e., articles with a focus on syntactic theory) stays remarkably constant across the five volumes. There are between 23 and 32 relevant papers

[3] However, even if this were not the case, it would not affect any claims regarding a decline or increase of algorithmicity as such; it would just not directly point to an inverse correlation with representationality.

per volume, with the number of articles without *any* example numbering environment hosting the definition or formulation of a building block varying between 5 (in 1989) and 11 (in 2014); cf. (5).[4]

(5) *Number of articles considered*:

a.	2014: 24	(10 without any numbered building block)
b.	2013: 23	(11 without any numbered building block)
c.	1994: 23	(6 with without any numbered building block)
d.	1993: 32	(8 without any numbered building block)
e.	1989: 24	(5 without any numbered building block)

I take it that the most straightforward account of the evidence in (4) is to assume that there has indeed been a decline in work focussed on the algorithmic perspective to syntax; and I also think that by far the simplest explanation for this development is the increase in work on representations. Other explanations for the decreasing ATR Ratio which might initially be conceivable do not strike me as likely to be on the right track.

Thus, let me first emphasize that whereas it is of course possible that the raw numbers in footnote 4 are affected by a few false positives and a few oversights, all the individual decisions of whether a numbered item contains algorithmic material or not have, in my experience, proven to be quite easy for all journal articles, and should in most cases qualify as uncontroversial.

Second, it does not look as though there has been a substantial change in the style of writing journal articles in the field of syntactic theory, such that building blocks would have come to be laid out more often in other contexts than example environments (e.g., as part of the main text).

Third, one might think that a recent development in minimalist syntax might be held responsible, in an indirect way: Following Chomsky (2005; 2013), a lot of recent minimalist work identifies as possible building blocks of syntax not only principles that are language-specific and principles that can directly be acquired from linguistic data, but also *third-factor* building blocks, i.e., principles not specific to the faculty of language (e.g., principles of efficient computation). Could the move to third-factor principles be responsible for the decline of the ATR ratio, given that these principles must be very general by their very nature? I think this is also unlikely: All types of factors, to the extent that they work as building blocks in individual grammars, can and need to be spelled out in sufficient detail so that their predictions can be checked; this is a hallmark of generative grammar. Thus, when, say, an extremely general language-independent economy principle is suggested as a building block that figures prominently in some analysis, it will be unavoidable, from an algorithmic perspective, to precisely state what consequences it has for syntactic analyses, and doing so requires a careful formulation of the building block's consequences (i.e., the theorem derived from it that is relevant for the case at hand) in syntactic terms.

---

[4] The raw data look as follows.

(i)
a. *2014*: 0/41, 4/121, 9/57, 0/59, 0/30, 3/30, 11/77, 2/74, 4/80, 0/33, 0/23, 0/15, 5/67, 0/51, 1/75, 1/22, 3/19, 0/26, 6/58, 3/63, 1/61, 1/61, 0/25, 0/26 = 54/1194.
b. *2013*: 2/89, 0/50, 0/26, 0/26, 1/14, 3/103, 6/85, 1/24, 0/16, 1/24, 0/18, 0/26, 3/114, 2/85, 9/38, 2/24, 0/19, 0/118, 3/59, 0/11, 0/34, 0/16, 3/21 = 36/1040.
c. *1994*: 5/76, 8/50, 19/156, 0/10, 0/18, 1/14, 5/60, 6/30, 1/51, 7/68, 1/18, 7/26, 1/61, 10/14, 0/63, 13/69, 0/29, 1/21, 0/17, 23/105, 6/100, 7/41, 0/8 = 121/1105.
d. *1993*: 16/46, 1/40, 5/44, 3/95, 3/17, 1/14, 0/14, 2/15, 6/105, 9/105, 22/80, 6/45, 1/16, 1/17, 0/10, 4/35, 7/15, 10/80, 2/35, 6/43, 0/16, 0/23, 2/19, 0/17, 6/135, 19/123, 0/0, 3/59, 5/45, 0/20, 2/11, 0/14 = 142/1353
e. *1989*: 18/96, 2/19, 1/6, 0/12, 0/24, 0/19, 9/66, 16/87, 10/35, 1/10, 2/16, 15/137, 16/84, 1/17, 2/11, 0/29, 2/6, 9/76, 18/85, 6/40, 7/29, 1/37, 0/9, 1/18 = 137/968

Since no alternative accounts of the decrease in ATR Ratios in *Linguistic Inquiry* suggest themselves in any obvious way, the claim that the field has to some extent moved away from an algorithmic perspective can thus count as independently supported.

To sum up, I have suggested that there has been a decline in research on syntactic theory from an algorithmic perspective which goes hand in hand with an increase in research on syntactic theory from a purely representational perspective; and I have argued that this claim can be substantiated by looking at the ATR Ratio in numbered examples of journal articles.

References

Belletti, Adriana, ed. (2004): *Structures and Beyond. The Cartography of Syntactic Structures*. Oxford University Press, Oxford.

Caha, Pavel (2009): The Nanosyntax of Case. PhD thesis, Tromsø University.

Chomsky, Noam (1957): *Syntactic Structures*. Mouton, The Hague and Paris.

Chomsky, Noam (1977): On Wh-Movement. In: P. Culicover, T. Wasow & A. Akmajian, eds., *Formal Syntax*. Academic Press, New York, pp. 71–132.

Chomsky, Noam (1981): *Lectures on Government and Binding*. Foris, Dordrecht.

Chomsky, Noam (1991): Some Notes on Economy of Derivation and Representation. In: R. Freidin, ed., *Principles and Parameters in Comparative Grammar*. MIT Press, Cambridge, Mass., pp. 417–454.

Chomsky, Noam (1995): *The Minimalist Program*. MIT Press, Cambridge, Mass.

Chomsky, Noam (2005): Three Factors in Language Design, *Linguistic Inquiry* 36, 1–22.

Chomsky, Noam (2013): Problems of Projection, *Lingua* 130, 33–49.

Cinque, Guglielmo (1999): *Adverbs and Functional Heads*. Oxford University Press, Oxford.

Fox, Danny (2000): *Economy and Semantic Interpretation*. MIT Press, Cambridge, Mass.

Heim, Irene & Angelika Kratzer (1998): *Semantics in Generative Grammar*. Blackwell, Oxford.

Höhle, Tilman (1986): Der Begriff 'Mittelfeld'. Anmkerungen über die Theorie der topologischen Felder. In: W. Weiss, H. E. Wiegand & M. Reis, eds., *Kontroversen, neue und alte. Band 3: Textlinguistik contra Stilistik? – Wortschatz und Wörterbuch – Grammatische oder pragmatische Organisation von Rede*. Niemeyer Verlag, pp. 329–340.

Kayne, Richard (1994): *The Antisymmetry of Syntax*. MIT Press, Cambridge, Mass.

Koopman, Hilda & Anna Szabolcsi (2000): *Verbal Complexes*. MIT Press, Cambridge, Mass.

May, Robert (1985): *Logical Form*. MIT Press, Cambridge, Mass.

Müller, Gereon (2014): *Syntactic Buffers*. *Linguistische ArbeitsBerichte* 91, Universität Leipzig.

Pesetsky, David (1982): Paths and Categories. PhD thesis, MIT, Cambridge, Mass.

Rizzi, Luigi (1997): The Fine Structure of the Left Periphery. In: L. Haegeman, ed., *Elements of Grammar*. Kluwer, Dordrecht.

Rizzi, Luigi (2004): Locality and Left Periphery. In: L. Rizzi, ed., *The Structure of CP and IP. The Cartography of Syntactic Structures, vol. 2*. Oxford University Press.

Starke, Michal (2009): Nanosyntax: A Short Primer to a New Approach to Language, *Nordlyd* 36, 1–6.

Stechow, Arnim von (1993): Die Aufgaben der Syntax. In: J. Jacobs, A. v. Stechow, W. Sternefeld & T. Vennemann, eds., *Syntax, vol. I*. De Gruyter, Berlin, pp. 1–88.

Stechow, Arnim von (1996): Against LF Pied-Piping, *Natural Language Semantics* 4, 57–110.

Stechow, Arnim von & Wolfgang Sternefeld (1988): *Bausteine syntaktischen Wissens*. Westdeutscher Verlag, Opladen.

Uriagereka, Juan (1999): Multiple Spell-Out. In: S. Epstein & N. Hornstein, eds., *Working Minimalism*. MIT Press, Cambridge, Mass., pp. 251–282.