Nanosyntax
A short primer to a new approach to language

Michal Starke
CASTL, University of Tromsø

Abstract

Nanosyntax is a novel approach to the architecture of language, designed to make (better) sense of the new empirical picture emerging from recent years of syntactic research. It is a large-scale project, addressing a wide array of issues, ranging from big issues such as the modularity of language, to fine details, such as the derivation of allomorphy in irregular patterns of given languages and its interaction with syntactic structures.

1. The atoms of language are smaller than we thought

The premise leading to the nanosyntactic project is very simple: Syntactic research has produced beautiful empirical generalisations over the last 30 years, and these generalisations have led to a profound change in the kind of mental representations ("syntactic structures") attributed to speakers. This profound shift has however remained within the empirical and notational domain, disconnected from syntactic theory itself: the theory used on the new structures is largely similar to the theory used 20 years ago (despite terminological changes). That theory is however not a good fit with the new results – the starting question was thus simple: what is the new empirical picture telling us? What do we **learn** from those beautiful generalisations?

A posteriori, the answer is surprisingly simple - though with deep consequences: the new syntactic structures are much larger, and growing by the day, and as a result, their ingredients (their terminal nodes) are getting much smaller. This turns out to contradict a fundamental tenet of the field: the deeply ingrained assumption that the ingredients of syntactic structure (the terminal nodes) are lexical items, "words" or "morphemes". The contradiction stems from the fact that orthodoxy views syntax as a way of arranging lexical items. But as syntactic structures grew, not only did their terminals become "much smaller", they became submorphemic - smaller than individual morphemes.

As a consequence, the terminals cannot be lexical items (morphemes or words), and hence syntax cannot be a device to arrange lexical items into structures. The field is thus in a position in which its fundamental assumptions are at odds with the results of its best research. We therefore need to reconsider the orthodoxy, questioning the very premise that syntax operates on lexical items. Nanosyntax is the result of doing that.

 Nordlyd 36.1, special issue on Nanosyntax, ed. Peter Svenonius, Gillian Ramchand, Michal Starke, and Knut Tarald Taraldsen, pp. 1–6. CASTL, Tromsø. http://www.ub.uit.no/baser/nordlyd/

2. Size differences

An immediate consequence of terminals being submorphemic is that many—perhaps most—morphemes will span several terminals. And therefore they will correspond to an entire subtree rather than corresponding to a terminal. This means that the lexicon contains subtrees, ie. syntactic trees, paired with phonological and conceptual information. Lexical entries will be minimally of the form < phonological information, syntactic tree, conceptual information >, and spellout becomes an operation matching the tree constructed by syntax to the (sub-)trees stored inside lexical entries (Starke 2002).

Conversely, syntax must now apply the same principles uniformly to features, "morphemes", "words" and "phrases" – they are all constituents in a rich syntactic tree, with features as terminals and all others as non-terminals.

This view makes sense of the results of syntactic research, i.e. the submorphemic nature of terminals: if lexical items correspond to entire constituents, it is evident that the ingredients of those constituents are smaller than lexical items. This view also offers a radically different view on the architecture of language: contrary to all other approaches, there is no lexicon feeding syntax. Syntax is an entirely pre-lexical system and the lexicon is a way of interpreting syntax (and mapping it onto other representations – such as conceptual representations and gestural (phonological) representations).

Empirically, this view also leads to many new and exciting research avenues. For instance, once lexical items are not confined into terminals anymore, they can be of different syntactic sizes - i.e. different lexical items may correspond to different amounts of syntactic structure. This basic observation gives us a new tool to understand various syntactic contrasts and has led to several research programs. One avenue that I have explored in this connection is that various sizes of lexical elements lead to different syntactic categories (eventive nouns are "bigger" than non-eventive nouns, verbs are bigger than nouns which are bigger than adjectives, etc). If feasible, this research line would lead to a non-parochial theory of syntactic categories – a long sought-after achievement. (See also Lundquist 2008 for discussions related to this line of thinking.)

The bulk of current research in nanosyntax focuses on two other consequences of "different sizes for different lexical items". First, morphemes now have internal syntactic structure, and hence we can capture in a principled way their varying behaviours, as well as the interaction between their ingredients and the rest of syntax. It was observed early on that nanosyntax allows for an elegant approach to syncretism (Starke 2002). If the matching between syntactic trees and trees stored in lexical items is not a perfect one-to-one mapping (i.e. a lexically stored tree matches a subtree of the syntax iff the two are exactly identical), but rather the matching allows subsets, then any lexical entry will map onto a range of syntactic trees, giving us a restricted theory of syncretism. The matching is thus along the lines of:

(1) A lexically stored tree matches a syntactic node iff the lexically stored tree contains the syntactic node

The formulation in (1) builds on the insight that any subtree of the lexically stored tree is also stored in the lexicon, and hence is a potential match for whatever syntax has built. As a consequence of (1), the lexical tree will always be a superset - proper or not - of the syntax it spells out. The matching principle in (1) is thus sometimes referred to informally as the Superset Principle.

The earliest empirical argument for (1) was based on the English suffix *-ed* (Starke 2002, Starke 2005). This suffix has both an active "reading" (*he danc-ed*, *he fold-ed the sheet*) and a passive "reading" (*The sheets were fold-ed*), with the two mapping onto a very different syntax. These two different syntaxes are however widely thought to be in a subset/superset relationship: a passive is a "crippled" version of the active, where the crippling is often expressed structurally (e.g. lack of "vP" in the passive). If the lexical representation of *-ed* corresponds to the bigger active structure, the matching principle in (1) will allow it to spell out the "crippled" subset structure of passives too. This thus derives the fact that *-ed* is syncretic between an active and a passive reading. In this early work, this approach to *-ed* was fleshed out and extended to the distinction between simple pasts, perfects, verbal passives and adjectival passives (both R-state and T-state). Ramchand (2006) is another early example of this logic, showing that the traditional puzzles about verb classes receive an elegant solution in terms of the superset (which Ramchand also calls 'underassociation') applying to a syntactic structure of the type [initiator [process [result]]].

3. Competition in size

The second aspect of "various sizes of lexical items" concerns the fact that lexical items will now compete between themselves in order to spell out whatever tree syntax has produced. The simplest situation is the case in which one lexical item is bigger than its competitors. In the simplest case, a lexical item can spell out the entire syntactic tree at once, and smaller competitors can spell out one daughter of the root node each. An example of this pattern is given by irregular plurals such as *mice*: the syntactic tree [N plural] is spelled out by two different lexical items in *elephant-s*, but by a single lexical item in *mice*. The lexical item *mice* stores the entire tree [N plural], but it has two competitors who could collaborate to spell out the same tree: *mouse*=N and *s*=plural, yielding **mouse-s*. The descriptive generalisation is always that:

(2) Theorem: biggest wins

This follows as a theorem from the current architecture of nanosyntax. Spellout is taken to be cyclic, with a spellout attempt after each merger operation. Each successful spellout overrides previous successful spellouts. Since merger is bottom-up, the biggest match will always override the smaller matches, and hence (2) follows as a theorem.

Another illustration of this situation is given by irregular past tenses: whereas *fold-ed* uses two different lexical items to spell out [V past], forms such as *went* or *flew* spell out the entire tree [V past] with a single lexical item. Again, the regular counterpart is blocked: **goed*, **flied*. And again, descriptively, the lexical item corresponding to the biggest subtree wins – a consequence of cyclic override.

The other major competition situation is where two different lexical items claim to be able to spell out a given syntactic node in one go. This is also illustrated by *mice*, *went*, *flew* etc. In the singular of 'mouse' for instance, there are two competitors for the syntactic structure 'N': *mouse* itself, and *mice* ([N plural]. This is because the matching principle (1) allows a subtree of *mice* to spellout the syntactic structure, and [N plural] does indeed contain a subtree N. Here the descriptive generalisation is again clear: the competitor which contains least unused material is the winner. As a convenient label I sometimes call this the MINIMISE JUNK principle – but it is in fact an instance of a better known and widespread principle: THE ELSEWHERE PRINCIPLE. The item with less "junk" is smaller, and hence matches fewer environments (given (1)) and is therefore more specialized, or "specific". Hence MINIMISE JUNK amounts to "the most specific wins", which is the elsewhere principle.

(3) At each cycle, if several lexical items match the root node, the candidate with least unused nodes wins (follows from THE ELSEWHERE PRINCIPLE).

These three ingredients (SUPERSET, CYCLIC OVERRIDE, ELSEWHERE PRINCIPLE) yield an elegant and surprisingly powerful theory of morphosyntax, syncretisms, allomorphies and generally of the syntax-lexicon correspondence. Starke (Starke 2002; Starke 2005) shows that it can also express the full range of irregular English forms alternating with *-ed*: the distribution of *-en*, *-ed* and opaque forms (e.g. *sit/sat/sat*) follows from their lexically stored syntactic trees and the two competition principles. There is no need for added power such as context-specific rules in the lexicon or in the morphology.

This general logic has since been shown to illuminate a variety of intricate morphosyntactic patterns. Caha (Caha 2009; Caha this volume) shows that the syncretisms in the case paradigms of nominal declensions follow from the nanosyntactic apparatus, analysing a wide range of languages. The logic of the superset, together with an innovative view of the underlying structure of case as cumulative privative features organised in a syntactic tree allows him to derive the existing syncretisms while at the same time ruling out the unattested syncretisms. This leads Caha to a new approach to case, which offers new insights into many case-related phenomena. Similarly, Taraldsen (to appear) shows that the size difference offers a promising solution to a traditional problem of morpho-syntax: the partial syncretisms of Bantu class markers, which have always resisted a principled account. Taraldsen successfully derives these syncretisms from the superset principle together with an articulated view of their underlying syntactic structure. A number of other lines of research have stemmed from the same logic and framework (Taraldsen and Medová 2007, Son and Svenonius 2008, Pantcheva this volume, etc) – too many to summarise here.

4. A new view of modularity and interfaces

Many more avenues of research are opened by nanosyntax, both within and beyond the size issue. Two prominent examples may bear mentioning. The phrasal spellout aspect of nanosyntax offers at long last an avenue into idioms. Confining lexical items into terminals is incompatible with the simplest fact about idioms: they are multiterminal expressions stored as such in the lexicon. As a result, there are no credible approaches to idioms within theories that restrict spellout to terminals. Nanosyntax expresses the multi-word and multi-terminal aspect of idioms directly – by simply storing the entire constituent (eg [VP= kick the bucket]) in a lexical entry. Of course, many issues about idioms remain unresolved (and often undescribed), but phrasal spellout opens the way confronting these issues productively.

Another example of a new direction afforded by nanosyntax and phrasal spellout is the issue of templatic effects in the morpho-phonological realm. Many templatic phenomena are cases in which a non-terminal node of morpho-syntax is targeted by a phonological restriction (e.g. the node must correspond to “two moras”, etc). In many such cases, the constituent undergoing the phonological templatic restriction is neither a “word” nor a “morpheme”, it is an intermediate node in the representation - a “phrasal” node. The question is how to express the correspondence between that constituent and the phonological template. This issue has remained a mystery, and is largely unapproachable in traditional theories. Nanosyntax however provides an answer: since entire trees are stored in the lexicon, a lexical entry will have no problem associating a phonological constituent (the template) with a syntactic phrase.

Phenomena such as idioms or phrasal templates can be handled by nanosyntax but not by traditional syntax and they thus constitute an important additional argument for the nanosyntax approach. Similar new considerations hold of the relationship between syntax and semantics – a topic of much ongoing interest among nanosyntax researchers.

5. Conclusion

Nanosyntax offers a radically new architecture of grammar: it departs from the concensus that "syntax projects from the lexicon". Syntax projects from single features and nothing else. Single features are merged together into the familiar binary branching trees, eventually attaining the size of a morpheme, a word and a phrase. Syntax doesn't build on morphemes, it builds morphemes.

There is a sense then, in which this syntax is "language-free": strictly language-related units such as morphemes or words are simply absent from syntax. Syntax is rather an abstract recursive grouping of formal features such as 'count', 'singular', 'definite', 'eventive', 'past', etc. with no knowledge of words or morhpemes. It is entirely conceivable then, that language is only one of several ways of interpreting this abstract structure – a result that may make sense of recent experimental results in cognitive science, though this is far beyond the scope of this note.

References

Caha, Pavel. 2009. *The Nanosyntax of Case*. Ph.D. thesis, University of Tromsø.

Caha, Pavel. this volume. Classical Armenian declension. pp. 77–112.

Lundquist, Björn. 2008. *Nominalizations and Participles in Swedish*. Ph.D. thesis, University of Tromsø.

Pantcheva, Marina. this volume. Directional expressions cross-linguistically: Nanosyntax and lexicalization. pp. 7–39.

Ramchand, Gillian. 2006. Direct and indirect causation in Hindi/Urdu. Handout. Workshop on Clitics, University of Konstanz, Germany.

Son, Minjeong and Peter Svenonius. 2008. Microparameters of cross-linguistic variation: Directed motion and resultatives. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*, edited by Natasha Abner and Jason Bishop, pp. 388–396. Cascadilla, Somerville, Ma.

Starke, Michal. 2002. The day syntax ate morphology. Class taught at the EGG summer school, Novi Sad.

Starke, Michal. 2005. Nanosyntax. Seminar taught at CASTL, University of Tromsø.

Taraldsen, Knut Tarald. to appear. The nanosyntax of Nguni noun class prefixes and concords. *Lingua* .

Taraldsen, Tarald and Lucie Medová. 2007. The Czech locative chameleon. In *Tromsø Working Papers on Language and Linguistics: Nordlyd 34.2,* Special issue on Space, Motion, and Result, edited by Monika Bašić, Marina Pantcheva, Minjeong Son, and Peter Svenonius, pp. 300–319. University of Tromsø, Tromsø. Available at http://www.ub.uit.no/baser/nordlyd/.