Morphosyntax of specific and non-specific indefinite markers

JAKUB DEKIER [iD]

Abstract

This paper proposes a nanosyntactic analysis of non-specific, specific unknown and specific known indefinite markers, as identified in Haspelmath (1997). The cross-linguistic analysis of syncretism between these three types of indefinite markers reveals that they form a particular hierarchy of structural containment, non-specific < specific unknown < specific known. Using the methodological tools of the nanosyntactic framework (Caha 2009; Starke 2009, 2011), it is possible to propose a syntactic model explaining the derivation of non-specific, specific unknown and specific known indefinite markers and the attested patterns of syncretism observed cross-linguistically in paradigms of the three types of markers. In consequence, the analysis shows that the analyzed indefinite markers constitute a syntactic domain which conforms to the *ABA generalization (Bobaljik 2012).

CORRESPONDING AUTHOR:
Jakub Dekier

Adam Mickiewicz University in Poznań, PL

dekier.jd@gmail.com

1 Introduction

Indefinite pronouns constitute a closed class of expressions, which includes a number of structurally and functionally similar sets of forms. While it is not possible to provide a single consistent definition encompassing all the attested functions of indefinite pronouns, from the perspective of structure, most items in this class are formed through the merger of two main parts, an indefinite marker and an ontological category stem.[1] As examples, take the English *some-* and *any-* series:

(1)   *English some- series*
         a.   PERSON – *some-body, some-one*
         b.   THING – *some-thing*
         c.   PLACE – *some-where, some-place*
         d.   MANNER – *some-how*
         e.   TIME – *some-time*

(2)   *English any- series*
         a.   PERSON – *any-body, any-one*
         b.   THING – *any-thing*
         c.   PLACE – *any-where, any-place*
         d.   TIME – *any-time*

An indefinite series is usually distinguishable by a distinct indefinite marker (morpheme), which determines its functional properties, and therefore, its usage and distribution. Particular items in a series differ with respect to the ontological category stem that the marker is merged with. The stem may take the form of a generic noun, wh-word or the word *one*, and its purpose is to specify the referential domain of the pronoun, for example, *some-thing* and *some-where* will refer to entities belonging to the categories of THING and PLACE, respectively.[2]

This paper discusses one of the two main elements comprising an indefinite pronoun, namely the indefinite marker. Specifically, the analysis concerns markers used in three particular indefinite functions: non-specific, specific unknown and specific known markers. Indefinite pronouns formed by these three markers are discussed in Haspelmath (1997: 37–52) in the following way. According to Haspelmath's study, non-specific pronouns indicate that the referent is not a particular entity (the speaker does not know or care if the referent exists; the referent may also be an unspecified item of a given group), while specific pronouns are used to describe a specific referent (the referent is a particular entity whose existence is presupposed). Additionally, specific pronouns can be either known or unknown. In the former case, the speaker is familiar with the identity of the referent, while in the latter, the actual identity of the entity the pronoun refers to remains unknown.

These three types of indefinite pronouns can also be seen on the map of indefinite pronoun functions put forward in Haspelmath (1997). The goal of Haspelmath's proposal was to show the variation in functional distribution of indefinite pronoun series found cross-linguistically. As a part of the typological analysis, it also is argued that an indefinite pronoun series may cover multiple functions, but only if those functions appear as contiguous elements on the map:
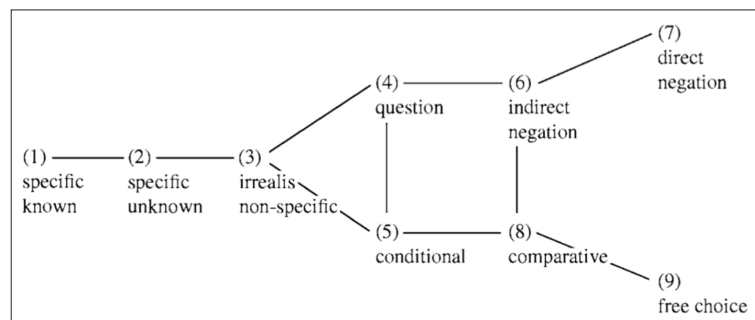

Figure 1 The implicational map for indefinite pronoun functions (Haspelmath 1997: 4).

As shown in *Figure 1*, the specific known, specific unknown and non-specific functions are placed next to one another, which means that just one indefinite pronoun series (or actually just one indefinite marker) may potentially cover all three functions. This is what can be observed in English:

---

1   In a number of cases, the indefinite marker and the categorical stem are not clearly separable morphemes, e.g. French *personne* 'nobody'.

2   Categorical stems represent what is known as ontological categories, which are a presumably closed class of functional nominals (cf. Baunaz & Lander 2019: 1–2; Cinque 2008: 18; Kayne 2005). The exact number of such categories is not known.

(3)   *English some-*
   a.   She wants to buy *some*-thing to read.   non-specific
   b.   She bought *some*-thing in that store. It was expensive.   specific unknown
   c.   I have *some*-thing to tell you. Guess what!   specific known

However, a possibility that I would like to explore in this analysis is that examples such as (3) do not show one indefinite structure in three context-dependent functions, but three separate syntactic entities lexicalized by a single phonological exponent (marker). As revealed in a cross-linguistic study of indefinite markers, the three adjacent functions (non-specific, specific unknown and specific known) may be lexically realized by one, two or three separate morphemes. When arranged in a paradigm based on the map of functions proposed in Haspelmath (1997) and their natural semantic compositionality (i.e. 1. non-specific, 2. specific unknown, 3. specific known), indefinite morphemes corresponding to the three functions show different patterns of syncretism. Moreover, out of the four patterns of syncretism possible in a three-item paradigm, only the ABA pattern is not attested in any of the studied languages (given the order in *Figure 1*). The absence of this pattern is in line not only with the predictions made in Haspelmath (1997: 76–82) but also a well-documented generalization concerning ordered sets (paradigms) of forms known as the *ABA generalization (Bobaljik 2012, see also Bobaljik 2007).[3]

All the attested patterns of syncretism and the absence of the ABA pattern can be explained through the use of methodological tools provided by Nanosyntax (Caha 2009, 2020, Starke 2009, 2018), which is a model of grammar oriented towards uncovering fine-grained syntactic structures. The nanosyntactic understanding of the phenomenon of syncretism leads us to put forward a claim that indefinite markers used in the non-specific, specific unknown and specific known functions correspond to a single universal syntactic hierarchy (sequence of features or rather sets of features). Elements F_1, F_2 and F_3 can be used to represent the order in which the layers of the hierarchy are derived as well as the levels of structural containment:[4]

(4)   Indefinite hierarchy – containment
   a. $[F_1] \Rightarrow$ non-specific marker
   b. $[[F_1]\ F_2] \Rightarrow$ specific unknown marker
   c. $[[[F_1]\ F_2]\ F_3] \Rightarrow$ specific known marker

Each of the three markers (morphemes) spells out a different subset of the hierarchy, which means that pieces of the syntactic structure corresponding to the three markers will be embedded into one another. In other words, the non-specific, specific unknown and specific known indefinite markers constitute phonological exponents of syntactic entities formed through feature cumulation.

On the basis of the studied data sample, it can be claimed that the sequence of features used to derive non-specific, specific unknown and specific known markers is cross-linguistically universal. Therefore, all languages can be predicted to follow this sequence of features for the derivation of the three indefinite marker types. At the same time, the evidence shows that the sequence may be lexicalized in different ways, which accounts for the observed differences between languages.

2 The three types of indefinite markers

The following section discusses the non-specific, specific unknown and specific known functions of indefinite markers and the differences between them on the basis of the typology proposed in Haspelmath (1997: 40–49). Additionally, the presented examples show the contrast between indefinite markers in English, a language where all three functions are morphologically represented by a single marker *some-*, and Russian, in which the three functions are marked with separate morphemes.

---

3   The lack of non-contiguous syncretic cells in a paradigm. See below.

4   The labels used here are primarily meant to represent the levels of embedding, i.e. how particular layers realize the subsets of the hierarchy. In this analysis I am not going to postulate the semantic content of F_1, F_2 and F_3.

First, consider indefinite markers used in the non-specific function. Indefinite pronouns formed on the basis of markers used in this function do not refer to a particular entity of the specified category. The existence of a referent is uncertain, not presupposed or impossible. The described entity may also be an unspecified item belonging to a given category or group (cf. Croft 1983; Haspelmath: 41–49):

(5)   *English* – non-specific indefinites
   a.   Mary wants to buy *something* for her sister. (She still doesn't know what to buy since she doesn't know what her sister likes).
   b.   On Saturday, they will go *somewhere*. (They still haven't decided where they want to go).
   c.   Bring me *something* to eat. (I don't care what you get for me).
   d.   If you ring the bell, *someone* may come. (Although, it is possible that we are completely alone here and nobody will come).
   e.   I wish I had *something* to write on. (Alas, I don't have any paper).
   f.   They will send *someone* from the company tomorrow. (We have no idea which employee will come).

In all the examples above, the grammatical or discourse contexts rule out the possibility that there is a particular entity which the indefinite pronoun may refer to. This means that only the non-specific reading is possible; the speaker either does not know (or care) if the referent actually exists, e.g. (5-a), or the referent itself is not an existing entity, e.g. (5-e). The referent may also be a random unknown member of a group, e.g. (5-f).

Non-specific indefinites will also appear in sentences containing modifiers or modals introducing uncertainty:

(6)   *English*
   a.   Apparently, *someone* (non-specific) is approaching.
   b.   He may go *somewhere* (non-specific) later.
   c.   She will probably buy *something* (non-specific).
   d.   Maybe *something* (non-specific) happened.

In Russian, the same non-specific function is expressed with the use of indefinite pronouns containing the indefinite marker *-nibud* (cf. Eremina 2012: 6–9; 72–77; Haspelmath 1997: 40):

(7)   *Russian* – non-specific indefinites
   a.   Ona pokupala   *što-nibud*   i   shla   domoj.
          She   bought.IMPERF what-INDEF and went.IMPERF home
          She bought *something* and went home. (usually)

   b.   Kažetsja, *kto-nibud'* približaetsja.
          seems   who-INDEF   approaches
          Apparently *someone* is approaching. (The speaker is not sure).

   c.   On skazal   mne vziat'   / chtoby ja vzial   *što-nibud*.
          He   tell.PAST me   take.INF / that   I   take-PAST what-INDEF
          He told me to take *something* (it didn't matter what).

   d.   On sprosil   nas, vstretili   li   my *kogo-nibud'* v   parke.
          He   ask-PAST us   meet-PAST whether we who-INDEF   in the park
          He asked us whether we met *anyone* (someone non-specific) in the park.

In contrast with non-specific markers, indefinite markers used in the specific function assert the existence of a particular referent. The speaker refers to an entity that exists within the frame established by the discourse. Consequently, since specific indefinite pronouns have a particular referent, they will appear in existential constructions (8-a) or contexts where the existence of the entity that the pronoun refers to is implied, for example, when an anaphoric pronoun is used (8-b). The non-specific reading is impossible in examples (8-a) and (8-b) (cf. Heringer 1969: 90; Karttunen 1976: 366, as cited in Haspelmath 1997: 41):

(8)   *English* – specific indefinites
a.   There is *something* (specific/*non-specific) that she wants to buy.
b.   She wants to buy *something* (specific/*non-specific) to read. Unfortunately, it is
expensive.

Furthermore, specific indefinite pronouns may be replaced with phrases denoting a particular
entity such as *a certain* + noun (Haspelmath 1997: 41):

(9)   *English*
a.   *Someone* (specific/*non-specific) broke into the house.
b.   *A certain person* (specific/*non-specific) broke into the house.

Specific indefinites are also often the only possible ones in certain realis contexts such as
perfective past or ongoing present (cf. Haspelmath 1997: 42) since the circumstances and
participants are fixed:

(10)   *English*
a.   *Someone* (specific/*non-specific) broke into the house. (He stole the TV).
b.   Look, *someone* (specific/*non-specific) is running. (He is wearing a blue shirt).

The contrast between non-specific and specific indefinites is clearly seen in Russian, where the
non-specific marker *-nibud* cannot be used when the referent is a specific entity. In such cases,
the specific marker *-to* has to be used:

(11)   *Russian* – non-specific vs. specific indefinites (Eremina 2012: 8–9)
a.   Ona pokupala   *što-nibud/*što-to*   i   shla   domoj.
She bought.IMPERF what-INDEF   and went.IMPERF home
She bought *something* (non-specific) and went home. (usually)

b.   Ona kupila   *što-to/*što-nibud* i   poshla   domoj.
she   bought.PERF what-INDEF   and went.PERF home
She bought *something* (specific) and went home (once).

In sentence (11-a), the action of buying was performed repetitively, and a different thing was
bought every time. The indefinite marker used in this sentence is *-nibud* and it denotes the non-
specificity of the referent. In contrast, when the action was performed only once and a specific
object was bought, the marker *-to* has to be used (11-b). Consider some other examples showing
that the non-specific marker *-nibud* cannot be used when the referent is perceived as specific:

(12)   *Russian* (Eremina 2012: 10–12, 72–77)
a.   Za   stenoj *kto-nibud/kto-to* smejalsia.
behind wall   who-INDEF   laugh-PAST
Someone was laughing behind the wall.

b.   Masha prigotovit *što-nibud/što-to* vkusnoje na uzhin.
Masha cook-FUT what-INDEF   delicious for dinner
Masha will cook *something* delicious for dinner (and she knows what it is going
to be, but the speaker does not).

Similarly, the specific marker *-to* will not be used when no particular referent is identified:

(13)   *Russian* (adapted form Eremina 2012: 44, 73–74; Haspelmath 1997: 40)
a.   Ty kupila   *što-to/što-nibud'* na uzhin?
You buy-PAST what-INDEF   for dinner?
Did you buy *anything* (something non-specific) for dinner?

b.   Kažetsja, *kto-to/kto-nibud'* približaetsja.
seems   who-INDEF   approaches
Apparently *someone* is approaching. (The speaker is not sure).

c.   On skazal   mne vziat'   / chtoby ja vzial   *što-to/što-nibud.*
He tell-PAST me   take-INF / that   I   take-PAST what-INDEF
He told me to take *something* (it didn't matter what).

    d.   On sprosil nas, vstretili li   my **kogo-to/kogo-nibud'* v parke.
        He ask-PAST us meet-PAST whether we who-INDEF   in the park
        He asked us whether we met *anyone* (someone non-specific) in the park.

It should however be noted that it is not impossible for indefinite pronouns to appear in the non-specific function in past or ongoing present contexts. Sentences describing habitual or repeated actions and sentences with universal quantifiers such as *every* or *each* may contain both non-specific and specific indefinites. The indefinite pronoun will be non-specific when the circumstances present a choice from a group of unspecified referents:

(14)   *English*
    a.   Every student is reading *something* (they are reading the same thing – specific).
    b.   Every student is reading *something* (they are reading different things – non-specific).

(15)   *English*
    a.   Everyday, *someone* would come and light the candles (a particular person would come – specific).
    b.   Everyday, *someone* would come and light the candles (could have been a different person each time – non-specific).

In *Russian*, the indefinite marker will have to match the intended interpretation:[5]

(16)   *Russian* (Eremina 2012: 31)
    a.   Každyj mal'čik budet rad esli vstretit *kogo-nibud'/*kogo-to* iz
        every boy be.FUT happy if meet.FUT who-INDEF   from
        svoix odnoklassnic.
        his girl-classmates
        Every boy will be glad if [he] will-meet *some* (someone) of his girl-classmates (it does not matter which one).

    b.   Kazhdyj prepodavatel' slyshal, chto *kogo-to/*kogo-nibud'* iz moix
        every teacher hear-PAST that who-INDEF   from my
        studentov vsegda vyzyvajut k dekanu.
        students always call-PRES-3RDPL(impers.) to dean
        Every teacher heard that *some (one)* of my students is always called before the dean (the same specific person is called every time).

Now consider one more example which again juxtaposes the two types (cf. Aloni & Port 2013; Haspelmath 1997: 41):

(17)   Non-specific and specific indefinites
    a.   Mary wants to marry *somebody* from the USA because she is American. She doesn't want to marry a foreigner.
    b.   Mary wants to marry *somebody* from the USA. They met on holiday in Mexico.

In example (17-a), Mary wants to marry an American person because she herself is American and apparently does not like the idea of marrying a foreigner. The speaker does not have a particular person in mind and does not presuppose that there is a specific individual that Mary

wants to marry. In contrast, (17-b) makes it obvious that there exists a particular person from the USA that Mary intends to marry. After all, Mary met that person on holiday in Mexico. For this reason, the interpretation of *somebody* in the second sentence is specific.

Indefinite markers used in the specific function shown in examples such as (8) and (12) can be described as specific unknown. This is because the exact identity of the referent remains unknown; the speaker does not know who or what the referent is. The term specific unknown is used to distinguish this function from another specific one, namely the specific known function (Haspemath 1997: 41–50):

(18)   *English* – specific known indefinites
      a.   I have *something* for you. (Try to guess what it is).
      b.   Mary has made *something* delicious for dinner. (I know you will love it).
      c.   I met *somebody* on the way. (It turned out to be a friend of mine).

In the examples above, not only is the referent a specific entity, but it is also familiar to the speaker, which means that they have information about its identity. For example, in (18-a), the speaker is talking about a particular gift, and they also know exactly what kind of gift it is. Similarly, in (18-b), the food made for dinner is a specific entity, and the speaker knows what it is (because they have seen it or heard about it). In general, indefinite pronouns are used in the specific known function when despite their knowledge of the referent, the speaker decided to withhold the information about its identity.[6]

The contrast between specific unknown and specific known indefinite pronouns can also be illustrated with examples from Russian. The specific marker *-to* will be used only when the speaker does not know the identity of the referent, and will be replaced by *koe-* in contexts where the identity of the referent is known to the speaker:

(19)   *Russian* – specific known indefinites (Eremina 2012: 9, 22–23)
      a.   Ja nashla *koe-što*   interesnoje v etoj knige.
            I   found   INDEF-what interesting in this book
            I found *something* interesting in this book.

      b.   Ja tebe   *koe-što*   prines. Smotri, kakaja dynia.
            I you-DAT INDEF-what brought. Look   which melon.
            I brought you *something*. Look at this melon.

As shown in the examples above, indefinite markers which appear in the non-specific, specific unknown and specific known functions have separate meanings and are used to describe different kinds of referents. The referential domains of the three indefinite functions do not overlap, which is especially evident in the examples from Russian, in which each function is expressed by a separate indefinite marker. This kind of data strongly indicates that each of the three indefinite functions corresponds to a distinct underlying syntactic structure. Subsequently, it can be argued that the structures that give rise to three separate indefinite markers in Russian can also be found in all languages in which the non-specific, specific unknown and specific known indefinite functions are attested, even if two or three of these functions are expressed by the same lexical item. English is one of the languages where only a single phonological exponent (marker) is used (*some-*) to represent all three of the indefinite functions. This means that in English, the indefinite markers (understood as morphemes used in particular functions) used in the non-specific, specific unknown and specific known functions are fully syncretic.

In Section 3, I will further explore the idea that the non-specific, specific unknown and specific known indefinite functions always correspond to distinct syntactic structures, regardless of the number of indefinite markers. I will attempt this by analyzing the patterns of syncretism found in the indefinite marker inventories of languages in a selected data sample.

---

6   The knowledge of the listener appears to be irrelevant. The collected data did not show any evidence of indefinite forms connected with the knowledge of a potential listener. The reason may be that the speaker often does not know what discursive information the listener actually has.

3 Indefinite markers – data

The analysis proposed in this paper is based on data from 45 languages: Basque, Bulgarian, Catalan, Czech, Dutch, English, Filipino, Finnish, French, Slovak, Georgian, German, Greek, Hausa, Hebrew, Hindi, Hungarian, Icelandic, Irish, Italian, Japanese, Kannada, Kazakh, Classical Greek, Korean, Latin, Latvian, Lithuanian, Lezgian, Maltese, Mandarin Chinese, Nanay, Ossetic, Persian, Polish, Portuguese, Quechua, Romanian, Russian, Serbian/Croatian, (Colombian) Spanish, Swahili, Swedish, Turkish, and Yakut. The data was analyzed with the aim of identifying the lexical forms of non-specific, specific unknown and specific known markers in each language.[7]

The majority of the data was taken from Haspelmath (1997), which constitutes a large-scale cross-linguistic analysis of indefinite pronouns.[8] Where possible, data concerning indefinite pronouns was also collected from other sources such as native speakers, grammar books and other linguistic literature. The language sample revealed a number of reoccurring patterns of syncretism between non-specific, specific unknown and specific known indefinite pronouns. The observed patterns constitute the basis of the analysis of the three types of markers proposed in this paper. Below, I provide examples of languages in which the particular patterns are attested.

3.1 Full syncretism (AAA pattern)

First, consider the data from English once more. In English, all three functions (non-specific, specific unknown and specific known) are represented by only one lexical item *some-*, which means that English non-specific, specific unknown and specific known markers are fully syncretic:

(20)　　*English*

    a.　　Mary wants to mary *some-one* from the USA. She hasn't   non-specific
        met the right man yet.

    b.　　There is *some-body* in the bathroom.   specific unknown

    c.　　I have *some-thing* to tell you. Guess what?   specific known

This pattern is quite common and is attested in a large number of languages, for example: Spanish, Latvian, Dutch, Icelandic, Bulgarian, Kazakh, Hungarian, Hindi, Maltese and Hebrew. Consider some examples from Polish, Japanese and Korean, in which the three types of markers are also fully syncretic:[9]

(21)　　*Polish – -ś* marker (native speakers)

    a.　　Przyprowadź mi　*kogo-ś*　　　　kto　　　zna　　angielski.
        bring.IMPV　I.DAT who.ACC-INDEF who.NOM knows English.ACC
        Bring me *someone*(non-specific) who knows English.

    b.　　*Kto-ś*　　　　　　jest　w　łazience.
        who.NOM-INDEF is　in bathroom.LOC
        Someone (specific unknown) is in the bathroom.

    c.　　*Kto-ś*　　　　　　dzwonił. Zgadnij　　kto.
        who.NOM-INDEF phoned　guess.IMPV who.NOM
        Someone (specific known) called. Guess who.

(22)　　*Japanese – -ka* marker (Haspelmath 1997: 312)

    a.　　*Dare-ka*　　ni　kiite　　　mimashou.
        who-INDEF DAT ask-CONV try-POL-HORT
        Let's ask *somebody* (non-specific).

---

7   A full description of the data can be found in the appendix.

8   It should be mentioned that in a few cases, the functional distribution of indefinite pronouns given in Haspelmath (1997) was not confirmed by native speakers and other written sources.

9   In Korean, the stem *mues* 'what' requires a linking morpheme *-i-*, e.g. *mues-i-nka* 'something', *mues-i-na* 'anything', *mues-i-tunci* 'anything'.

b. *Dare-ka*   kare denwa atta   kedo, dare kara da   ka
who-INDEF from phone be.PAST though who from be.PRES Q
wakaranai
know-Neg-PRES
Somebody (specific unknown) called, – I don't know who.

c. *Dare-ka*   kare denwa atta   kedo, dare kara da   ka
who-INDEF from phone be.PAST though who from be.PRES Q
atete   goran
figure.out-CONV try-IMPV
Somebody (specific known) called, – Guess who.

(23)   *Korean – -nka* marker (Haspelmath 1997: 314–315)

     a.   Molu-myen,   *nwukwu-nka-eykey* mul-ela.
ignorant-CONV who-INDEF-DAT   ask-IMPV
If you don't know, ask *somebody* (non-specific).

     b.   Ku sonyen-un *mues-inka-lul*   po-ass-ta.
this boy-NOM what-INDEF-ACC see-PAST-DECL
The boy saw *something* (specific unknown).

     c.   *nwukwu-inka-ka*   kel-ess-ta.
who-INDEF-NOM call-PAST-DECL
Somebody (specific known) called.

3.2 No syncretism (ABC)

As mentioned in Section 2, Russian is a language with no syncretism between the three markers, which are realized as separate lexical forms (see examples in Section 2).[10] Therefore, the paradigm of the three indefinite markers in Russian is as follows:[11]

(24)   *Russian*

     a.   što-*nibud*   something non-specific
     b.   što-*to*   something specific unknown
     c.   *koe*-što   something specific known

Another language where no syncretism is observed is Lithuanian (cf. Haspelmath 1997: 276; Kozhanov 2015; Pilka 1984: 97,29). The three markers are *-nors, kaž-* and *kai-*:

(25)   *Lithuanian* (Haspelmath 1997: 276; Kozhanov 2015: 2)

     a.   Jei tu   *ką-nors*   matai, pasaky-k man.
if   you what-INDEF see   tell-IMPV to:me
If you see *anything* (something non-specific), tell me.

     b.   Jei tu   *kaž-ką*   matai, pasaky-k man.
if   you INDEF-what see   tell-IMPV to:me
If you see *something* (specific unknown), tell me.

     c.   Turiu *kai-ką*   tiktai tau   vienai pasakyti.
I:have INDEF-what only   to:you alone to:say
I've got *something* (specific-known) to say that's for your ears alone.

Just as Russian, Lithuanian uses three separate markers for the non-specific, specific unknown and specific known indefinite functions:

(26)   *Lithuanian*

     a.   kas-*nors*   something non-specific
     b.   *kaž*-kas   something specific unknown
     c.   *kai*-kas   something specific known

10   The *-libo* marker may be used as a formal version of *-nibud* and there is no difference between the two (cf. Eremina 2012: 72). For this reason I will not treat *-libo* forms as a separate indefinite series. There is also the *ne*-marker, which, as far as my data suggests, is rarely used. It can be considered formal or emphatic and is similar to the standard specific *-to* series. Additionally, pronouns in this series are always used in the nominative.

11   In the example paradigms, I will use forms corresponding to the category of THING, for example, the generic-noun-based *some-thing* in English or wh-based *što-to* in Russian.

3.3 ABB syncretism

Another pattern of syncretism is observed in Ossetic, Yakut, Georgian and Nanay. These languages lexically distinguish only specific and nonspecific markers. In other words, the specific known and unknown markers are syncretic to the exclusion of the non-specific marker. First consider the data from Ossetic (Haspelmath 1997: 281; Kulaev 1958: 52), in which the morpheme *-dær* is used in the specific unknown/known functions and *is-* appears in the non-specific function:[12]

(27)   *Ossetic* (Haspelmath 1997: 281; Kulaev 1958: 52)
    a.   *Cy-dær*   mæ   qygdær-y.
          what-INDEF   me   bother-PRES.3SG
          Something (specific unknown/known) bothers me.

    b.   Mænæm dær ma *is-ty*   ratt.
          I:DAT   also PT INDEF-what give(IMPV)
          Give me *something* (non-specific), too.

The same kind of pattern is attested in Yakut (Afanas'ev and Xaritonov 1968; Haspelmath 1997: 290; Ubrjatova 1982). The forms used in Yakut are *-ere* (specific unknown/known) and *-eme* (non-specific):

(28)   *Yakut* (Haspelmath 1997: 290)
    a.   *Kim-ere*   eji-exe   kiir-e   syld'y-byt-a.
          who-INDEF   you-DAT   enter-CONV go-PERF-3SG
          Someone (specific unknown/known) has come to you.

    b.   Xojut   *kim-inen-eme*   *tug-u-eme*   yyt-ya-m.
          afterwards who-INSTR-INDEF what-ACC-INDEF send-FUT-1SG
          Afterwards I'll send *something* (non-specific) with *someone* (non-specific).

Georgian is another language where markers used in the specific known and unknown functions are syncretic (Haspelmath 1997: 304; Sharahsendize 2018). The morpheme *-γac* appears in the specific unknown/known functions, while *-me* is used in the non-specific function:

(29)   *Georgian* (Haspelmath 1997: 304)
    a.   Es   c'igni *sad-γac*   v-išove.
          this book where-INDEF 1SG-found
          I found this book *somewhere* (I could say where – specific known).

    b.   Movida *vi-γac*   rusi.
          came   who-INDEF Russian
           Some Russian person has come (I don't know him/her – specific unknown).

    c.   Dauaxet   *vi-s-me*!
          call:IMPV:PL who-DAT-INDEF
          Call *somebody* (non-specific)!

The ABB pattern can also be observed in Nanay. The morpheme *-daa* is used in the non-specific function, while *-nuu* appears only in specific contexts. It should however be noted that the data on this language presented in Haspelmath (1997) are described as incomplete. The author did not have detailed information about the specific functions (known/unknown). I will therefore omit Nanay in the summary below:

(30)   *Nanay* (Haspelmath 1997: 67–68; Onenko 1986)
    a.   *Uj-nuu*   žook-či   laŋ ži-či-ni.
          who-INDEF   house-DIR near go-PAST-3SG
          Someone (specific) went up to the house.

    b.   Ñoambani *xajla-nuu*   bajtalto-j-či.
          they   what-INDEF accuse-PRES-3PL
          They are accusing him of *something* (specific).

---

12   When combined with *is-*, *cy* 'what' chnages into *ty*. This appears to be a phonological change. Other items in the series remain unchanged, for example *kæm* 'where', *is-kæm* 'somewhere' and *kæd* 'when', *is-kæd* 'sometime'.

c. *Xaj-daa*   osii-daa osi-žara.
what-INDEF [?]   happen-FUT.3SG
Something (non-specific) may happen .

d. Sajna, *xaj-daa* žaka o-či-ni   bižere.
probably what-INDEF thing happen-PAST-3SG apparently
Probably *something* (non-specific) has happened.

To summarize, the specific unknown and specific known markers are syncretic in Ossetic, Yakut and Georgian:

(31)   *Ossetic*
    a.   *is*-ty   something non-specific
    b.   cy-*dær*   something specific unknown
    c.   cy-*dær*   something specific known

(32)   *Yakut*
    a.   tuox-*eme*   something non-specific
    b.   tuox-*ere*   something specific unknown
    c.   tuox-*ere*   something specific known

(33)   *Georgian*
    a.   ra-*me*   something non-specific
    b.   ra-*γac*   something specific unknown
    c.   ra-*γac*   something specific known

3.4 AAB syncretism

Latin represents a pattern, where the non-specific and specific unknown markers are syncretic to the exclusion of the specific known one. The non-specific and specific unknown functions are represented by *ali-* (Haspelmath 1997: 254; The Bible: Acts 3: 5, Luke 8: 46, Mark 9: 38):

(34)   *Latin*
    a.   Tetigit   me *ali-quis.*
          touched   me INDEF-who
          Somebody (specific unknown) hath touched me (for I perceive that virtue is gone out of me).

    b.   At ille intendebat in eos, sperans se *ali-quid*   accepturum
          but that gave.heed in them hoping self INDEF-what accept:FUT
          ab   eis.
          from them
          And he gave heed unto them, expecting to receive *something* (non-specific) of them.

The specific known function is expressed with the morpheme *-dam*:

(35)   *Latin*
    Magister, vidimus *quem-dam* in nomine tuo ejicientem daemonia.
    Master   we:saw who-INDEF in name   your casting.out devils
    Master, we saw *someone* (specific known) casting out devils in thy name.

Therefore, the lexical exponents used in Latin are as follows:

(36)   *Latin*
    a.   *ali*-quid   something non-specific
    b.   *ali*-quid   something specific unknown
    c.   quid-*dam*   something specific known

3.5 Syncretism – summary

As shown in this section, the non-specific, specific unknown and specific known indefinite functions can correspond to a varying number of lexical items, from just one to as many as

three. In other words, the non-specific, specific unknown and specific known indefinite markers (understood as morphemes used in particular indefinite functions) can be syncretic. *Table 1* shows the patterns of syncretism attested in the studied language sample:

	non-specific	**specific unknown**	**specific known**	**pattern**
English	some-	some-	some-	AAA
Yakut	-eme	-ere	-ere	ABB
Latin	ali-	ali-	-dam	AAB
Russian	-nibud	-to	koe-	ABC
unattested				ABA

The data shown in *Table 1* are arranged according to the map of indefinite functions (see *Figure 1*.), where the non-specific, specific unknown and specific known functions form a sequence of adjacent items. This ordering stems from a generalization made in Haspelmath (1997: 4), which says that a series of indefinite pronouns may express multiple indefinite functions, but only if those functions are adjacent on the map. In the context of syncretism, this generalization means that syncretism between indefinite markers should target only adjacent functions on Haspelmath's map. This claim seems to be confirmed, since there are no attested cases in which the specific known marker is sycretic with the non-specific markers to the exclusion of the specific unknown marker.

The prediction that syncretism has to target adjacent items in a sequence does not however apply solely to indefinite markers and connects to a broader generalization known as *ABA. As argued in Bobaljik (2012), the ABA pattern will not appear in ordered sequences (paradigms) where items form a hierarchy of structural containment. The absence of the ABA pattern is exactly what we notice when indefinite marker forms are arranged in a paradigm according to the ordering found on Haspelmath's map of indefinite functions. Whenever syncretism between indefinite markers is observed, it always targets contiguous cells in the proposed paradigm. Hence, if the *ABA generalization is correct, the patterns shown in *Table 1* lead us back to the initial proposal of this paper, namely that features corresponding to the non-specific, specific unknown and specific known indefinite markers form a cross-linguistically universal hierarchy of syntactic structures that are syntactically contained in one another. At this point, it should however be mentioned that the analysis of syncretism reveals only the relative order of elements in a sequence and does not inform us about the direction in which a syntactic hierarchy grows. This means that two orders of derivation should be considered:

(37)   Indefinite hierarchy – the two possible orders
    a.   non-specific < specific unknown < specific known
    b.   specific known < specific unknown < non-specific

While the specific unknown structure will inevitably be in the middle of the hierarchy, syncretism does not show which marker should correspond to the smallest layer of the sequence (F_1). A criterion that could be applied to establish the least and the most complex items in the hierarchy is morphological complexity. This means that we could expect that more complex indefinite structures will be phonologically realized as a greater number of morphemes, or that we will observe morpheme stacking. The morphological complexity criterion is however not useful in the case of indefinite markers. In the analyzed data sample, there are no cases of marker stacking or examples that can clearly indicate what types of markers are structurally more complex than others.[13]

---

13   This may be due to the fact that indefinite markers originate from a wide variety of particles, functional words and phrases. For example, in Russian, the non-specific marker *-nibud* comes from the phrase *ni budi* 'it may be', the specific unknown marker *-to* is related to the demonstrative *to* 'that' and the specific known marker *koe-* appears to be the neuter form of *koj* 'which'. The morphological structure of indefinite markers does not reveal any significant patterns cross-linguistically or within languages (apart from syncretism).

I will therefore use another criterion to establish which marker corresponds to the smallest structure in the hierarchy, namely functional compositionality. The idea is that the three types of markers can be arranged in a sequence on the basis of their growing functional complexity. Non-specific markers can be considered to correspond to the simplest syntactic structure, since they only introduce indefinite entities but lack other properties such as specificity or knowledge of the speaker. In contrast, specific unknown markers introduce an additional property, specificity of the referent. Lastly, specific known markers have yet another additional property, knowledge of the speaker. Thus, the functional properties of the three marker types reveal the relative complexity of their underlying syntactic structures; non-specific markers correspond to the smallest subset of the hierarchy, and specific known markers spell out the whole hierarchy.[14] The ordering proposed on the basis of the functional complexity of non-specific, specific unknown and specific known indefinite markers not only matches the relative order of the markers established on the basis of syncretism but also suggests that (37-a) is the correct ordering for the proposed hierarchy. For this reason, I will adopt (37-a) in the analysis.

The observed patterns of syncretism and the cross-linguistic absence of the ABA pattern, as predicted by the *ABA generalization, constitute strong evidence to support the claim that non-specific, specific unknown and specific known indefinite markers correspond to subsets of a structural containment hierarchy. However, it remains to be explained how exactly the hierarchy is lexicalized so that it gives rise to the observed patterns of syncretism. A clear and coherent answer to these questions follows from the methodology provided by Nanosyntax. Mechanisms introduced by the nanosyntactic framework, such as cyclic phrasal spellout and spellout-driven movement, allow us to explain the derivation of indefinite markers and the emergent patterns of syncretism.

4 Nanosyntax

Nanosyntax (Baunaz et al. 2018, Baunaz & Lander 2018b; Caha 2009, 2020; Starke 2009, 2011; see also Cardinaletti & Starke 1999) is an approach to the analysis of grammar which has its roots in the cartographic tradition (e.g. Cinque 1999; Cinque & Rizzi 2008; Rizzi 1997). The nanosyntactic framework is oriented towards studying syntactic representations to uncover the fine-grained structure of grammar.

Nanosyntax inherited some of its most fundamental claims from the cartographic approach. These core principles are as follows:

1. one feature-one head (OFOH) (Cinque & Rizzi 2008: 50, Kayne 2005)

2. the universality of the functional sequence (Baunaz & Lander 2018b: 16–21; Cinque & Rizzi 2008: 45; Starke 2001)

In line with the OFOH maxim, each projecting head (a terminal node) may contain only a single syntactic feature. In consequence, there is no feature bundling and features become the basic building blocks that syntax uses to form structures. In other words, every syntactic structure constitutes a particular hierarchy of features, for example:

(38)

$$
\begin{array}{c}
F_3P \\
\diagup \diagdown \\
F_3 \quad F_2P \\
\diagup \diagdown \\
F_2 \quad F_1P \\
\mid \\
F_1
\end{array}
$$
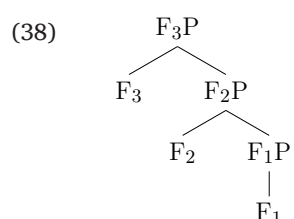
The fact that each head contains only a single feature makes terminal nodes smaller than words or even morphemes (they become submorphemic) (Starke 2018). One morpheme will often spell out multiple terminal nodes, for instance the English pronoun *he* will spell out at least four heads, since it phonologically represents case, person, number and gender features (among others). For this reason, it is argued that to spell out multiple heads as a single lexical item,

---

14 I use the terms "specificity" and "knowledge of the speaker" to establish the differences between the discussed indefinite markers and analyze the levels of structural embedding.

spellout has to target phrasal nodes rather than terminals. In other words, syntactic structures are are spelled out as constituents. The phrasal spellout mechanism is often considered to be one of the most important features of Nanosyntax.

Another immediate consequence of single-feature heads is the elimination of the boundary between morphology (word-building) and syntax as two separate modules of language. Since syntax operates only on submorphemic features, then all structure building, from morphemes and words to whole sentences, can take place within it. Thus, morphology, understood as a word-building mechanism separate from syntax, is no longer necessary and becomes a part of syntax.

Regarding fundamental claim number two, i.e. the order according to which features are merged into syntactic structures (also known as the functional sequence or the fseq), it is considered to be cross-linguistically universal. In other words, languages always merge features in the same relative order. In consequence, linguistic variation will stem from differences in how languages lexicalize the functional sequence (spellout), rather than from differences on the level of syntax (cf. Baunaz & Lander 2018a).[15]

4.1 The nanosyntactic lexicon and spellout (Starke 2009, 2014, 2018)

Due to the fact that terminal nodes in Nanosyntax contain only one feature, lexical items will not correspond to terminal nodes but to phrasal constituents. In consequence, the nanosyntactic lexicon constitutes an organized list of well-formed syntactic structures that a particular speaker is familiar with (Starke 2014: 1–2). Each piece of structure stored in the lexicon corresponds to a particular phonological exponent, for example:

(39) Lexical entry

$$/\alpha/ \Leftrightarrow F_1P$$
$$|$$
$$F_1$$

When spellout is triggered, it will involve accessing the lexicon in search for a lexical entry (a lexically-stored piece of structure) matching the structure built in syntax. If a matching lexical entry is found, its corresponding phonological exponent can be inserted into that structure (a syntactic constituent). This kind of spellout system makes Nanosyntax a late-insertion model of grammar (cf. Caha 2020: 1). No lexical information is carried by features or provided in any way before spellout takes place.

The Nanosyntactic lexicalization process in not only phrasal but also cyclic, in the sense that spellout is triggered each time a new feature is merged. In other words, with every feature merge, spellout will immediately access the lexicon in an attempt to find a lexical entry that matches the derived structure. Obtaining a matching entry will provide a phonological exponent for the phrase derived by the last merge operation. The outcome of the previous lexicalization cycle, i.e an exponent corresponding to a subset of the current structure, will be overriden by the newly inserted exponent.[16]
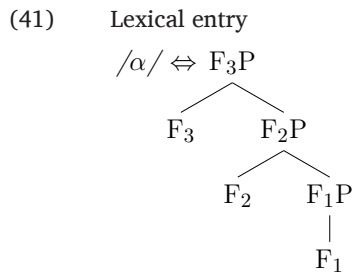
Now consider the following examples which illustrate the basic principles of the nanosyntactic spellout system:

(40)

$$
\begin{array}{c}
F_3P \\
\diagdown \\
F_3 \quad F_2P \\
\diagdown \\
F_2 \quad F_1P \\
| \\
F_1
\end{array}
$$

---
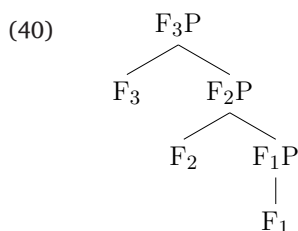
[15] It may also be argued that linguistic variation is caused by language-specific differences in the number of features in the fseq. According to this proposal, the number of features in the fseq is not the same cross-linguistically; only the relative order of those has to be constant. Arguably, this idea may be supported by languages in which we see the total absence of any traces of certain grammatical features, for example the neuter gender in Italian (nouns in Italian can only be masculine or feminine).

[16] This principle is sometimes referred to as the principle of cyclic override (Wiland 2019: 11).

To lexicalize the structure in (40), the spellout mechanism has to access the lexicon and find a lexically-stored tree that matches the derived structure, for example:
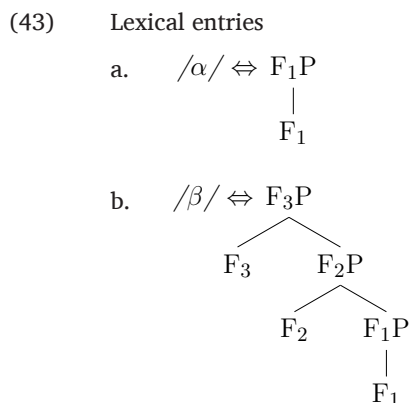
(41)   Lexical entry

$$/\alpha/ \Leftrightarrow F_3P$$

with structure: F_3P branches into F_3 and F_2P; F_2P branches into F_2 and F_1P; F_1P dominates F_1.

Once a matching lexical entry is found, its corresponding phonological exponent is inserted into the phrasal node projected by the last successfully merged feature (F_3P), and the system can proceed to merge the next feature in accordance with the fseq. The merger of a new feature will trigger another spellout cycle.

Matching between syntax and the lexicon is regulated by the Superset Principle. If no lexical entry is an exact match for a structure, it can still be spelled out with an overspecified lexical entry. In consequence, an exact match is not always necessary to spell out a structure. The Superset Principle states that:

(42)   The Superset Principle (Caha 2009; Starke 2009: 3)
   "A lexical item matches a syntactic node if it is a superset of that node."

This can be illustrated with the following example where two lexical entries are available:
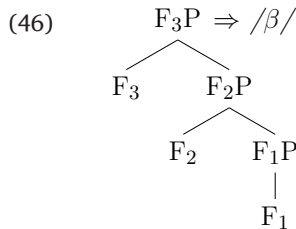
(43)   Lexical entries

   a. $/\alpha/ \Leftrightarrow F_1P$ dominating F_1

   b. $/\beta/ \Leftrightarrow F_3P$; F_3P branches into F_3 and F_2P; F_2P branches into F_2 and F_1P; F_1P dominates F_1

When F_1 is merged, lexical access is triggered and (43-a) becomes the matching entry. The phonological exponent α can be inserted:

(44) $F_1P \Rightarrow /\alpha/$ with F_1P dominating F_1
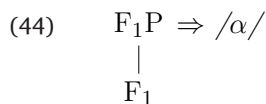
The next merge adds F_2 to the structure and a new spellout cycle begins. At this point, even though the lexicon contains no entry corresponding to the sequence [F_1, F_2], the Superset Principle will allow (43-b) to spell it out because the tree in (43-b) contains the derived structure. Thus, β will be inserted as the phonological exponent of [F_2 [F_1]], which will also overwrite the result of the previous lexicalization cycle:

(45) $F_2P \Rightarrow /\beta/$; F_2P branches into F_2 and F_1P; F_1P dominates F_1
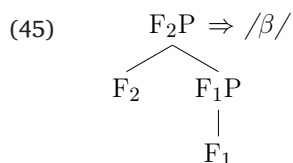
The lexicalization of [F_1,F_2,F_3] is again straightforward, since (43-b) is an exact match for the sequence:

(46)

$$F_3P \Rightarrow /\beta/$$

with children F_3 and F_2P; F_2P with children F_2 and F_1P; F_1P dominating F_1.
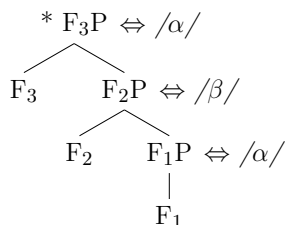
A question that naturally arises at this point concerns the spellout of F_1P with (43-b). After all, according to the Superset Principle, (43-b) is also a proper match for F_1P, since F_1P constitutes a subset of the tree in (43-b). This problem is solved by the second rule that governs lexical insertion in the nanosyntactic framework, namely the Elsewhere Principle:[17]

(47)   The Elsewhere Principle (Starke 2009: 4)
        "If several lexical items match a syntactic node, insert the entry with the fewest features unspecified for that node."
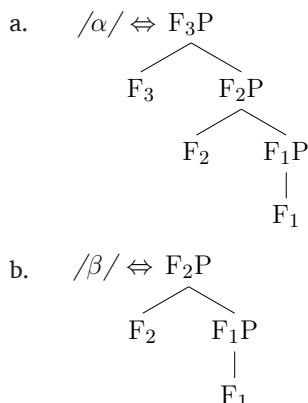
The Elsewhere Principle specifies that the closest match will always win in case of a competition between multiple lexical entries. For this reason, (43-b) will never be selected over (43-a) as the matching lexical entry for F_1P, as the latter contains fewer superfluous features.

The Superset Principle and the Elsewhere Principle explain syncretism and the *ABA generalization (Bobaljik 2007, Bobaljik 2012). The Superset Principle makes it possible for a single entry to spell out multiple structures that exist in a containment relation, which gives rise to syncretism. At the same time, the Elsewhere Principle constrains lexical insertion, so that more specific lexical entries will always win against ones containing more superfluous features. This means that a single lexical entry will not be used to spell out two non-contiguous layers of a syntactic hierarchy. In the nanosyntactic system, the following spellout results for the derivation of $[F_1, F_2, F_3]$ will be illicit (given the entries in 49):

(48)   *ABA as a consequence of phrasal spellout

$$* F_3P \Leftrightarrow /\alpha/$$

with children F_3 and $F_2P \Leftrightarrow /\beta/$; F_2P with children F_2 and $F_1P \Leftrightarrow /\alpha/$; F_1P dominating F_1.

(49)   Lexical entries
        a. $/\alpha/ \Leftrightarrow F_3P$

with children F_3 and F_2P; F_2P with children F_2 and F_1P; F_1P dominating F_1.

        b. $/\beta/ \Leftrightarrow F_2P$

with children F_2 and F_1P; F_1P dominating F_1.

## 4.2 Spellout-driven movement and subderivation (Starke 2018; Wiland 2019)

The Superset Principle makes it possible to lexicalize structures that do not have exact matches in the lexicon (as they can be spelled out with overspecified lexical entries), however, there may be cases where none of the available lexical entries match the derived piece of syntactic structure.

---

17   Here I refer specifically to the nanosyntactic Superset Principle, also known as the Superset Condition. See also Kiparsky (1973).

This presents a major problem for the lexicalisation mechanism (as described so far), assuming that the nanosyntactic spellout mechanism may not leave a feature without a phonological exponent.[18] In other words, all features have to be spelled out in each spellout cycle.

The nanosyntactic answer to this issue, and another important feature of the nanosyntactic spellout system, is a mechanism known as spellout-driven movement (cf. Caha 2011; Starke 2018). Whenever lexicalization is impossible (due to the lack of a matching lexical entry), the need to spell out the structure triggers syntactic movement to obtain a lexicalizable tree geometry and save the derivation from crashing. Below, I discuss an example of spellout-driven movement that is relevant to the analysis presented in this paper.

Consider a new example derivation where F_3 has just been added to the structure, triggering spellout. Additionally, assume that there is no entry in the lexicon that could be used to spell out the created sequence. As seen below, in the previous lexicalization cycle α was inserted as the phonological exponent of F_2P:

(50)

$$F_3P \Rightarrow \text{?}$$
$$F_3 \quad F_2P \Rightarrow /\alpha/$$
$$F_2 \quad F_1P$$
$$F_1$$
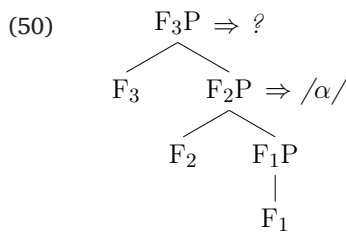
Now assume that the lexicon includes a lexical entry containing F_3, which however cannot be used because it does not match the sequence $[F_1, F_2, F_3]$. The entry matches a constituent containing only F_3:

(51)   Lexical entry

$$F_3P \Leftrightarrow /\beta/$$
$$F_3$$
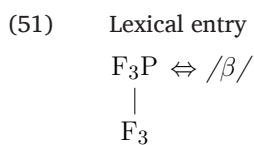
To use this entry, it is necessary to obtain a tree geometry, where F_3 forms a separate constituent. This can be achieved by a roll-up movement that evacuates the complement of F_3 to a position above it:[19]

(52)   Spellout-driven movement

$$F_3P$$
$$F_2P \qquad F_3P$$
$$F_2 \quad F_1P \quad F_3 \qquad \dots$$
$$F_1P$$
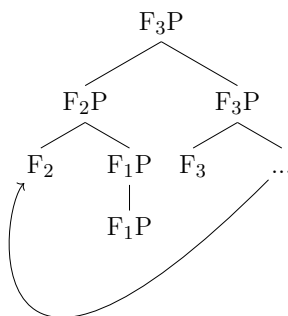
The result of the roll-up movement is a tree in which the sister of F_2P can be spelled out as $/\beta/$ in line with (51). This means that F_3P will now be lexicalized as a separate morpheme:

(53)

$$F_3P$$
$$/\alpha/ \Leftarrow F_2P \qquad F_3P \Rightarrow /\beta/$$
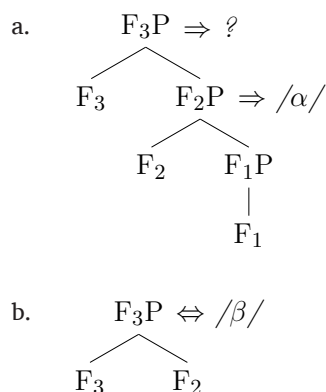$$F_2 \quad F_1P \quad F_3$$
$$F_1P$$

---
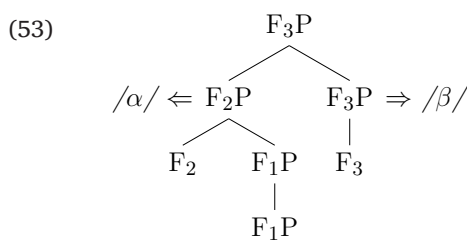
18   This rule is known as the Exhaustive Lexicalization Principle (Fábregas 2007).

19   This movement conforms to the Extension Condition (Chomsky 1993) and creates a non-projecting sister mode to the node that is to be spelled out. See also Pantcheva (2011: 135) and Wiland (2019: 13–14).

The spellout mechanism will always attempt to remerge a piece of structure if a matching lexical entry is not found (spec-to-spec movement and then roll-up movement).[20] In cases where movement cannot produce a lexicalizable structure, the system will backtrack to the previous merge cycle (undo the last merge operation) and apply transformations at that point. This operation, i.e. backtracking, may be attempted multiple times if necessary. However, if all of these steps (movement and movement after backtracking) fail to provide a structure in which all features can be spelled out, there is still one more option available, namely subderivation (Starke 2018).[21] This last-resort operation spawns a parallel derivation in order to construct a lexicalizable constituent containing the feature that the system is trying to spell out. The subderived hierarchy will subsequently be spelled out and integrated into the main structure as a complex left branch. Consider the following set of examples illustrating the process. Assume that it is impossible to spell out F_3 through movement (and backtracking) in (54-a) and the lexicon contains an entry such as (54-b):

(54) Lexical entries

a.

$$F_3P \Rightarrow \text{ }?$$

$$F_3 \qquad F_2P \Rightarrow /\alpha/$$

$$F_2 \qquad F_1P$$

$$F_1$$

b.

$$F_3P \Leftrightarrow /\beta/$$

$$F_3 \qquad F_2$$

Since it is not possible to save the derivation in any other way, the syntactic system is forced to form a subderivation. Because derivations begin with a merge, the parallel structure will have two features at the bottom: a copy of the last succesfully merged feature from the main sequence and F_3, which the system wants to lexicalize. It should however be noted that there is no consensus at the moment regarding the first feature that has to be merged at the bottom of a subderivation. In the presented analysis, I will follow Caha et al. (2019), which states that the main sequence and the parallel derivation will overlap. The last succesfully merged feature of the main derivation will appear at the bottom of the subderivation together with the feature that the system is trying to lexicalize:

(55) The subderivation (a) and the main spine (b)

a.

$$F_3P$$

$$F_3 \qquad F_2$$

b.

$$F_2P \Rightarrow /\alpha/$$

$$F_2 \qquad F_1P$$

$$F_1$$
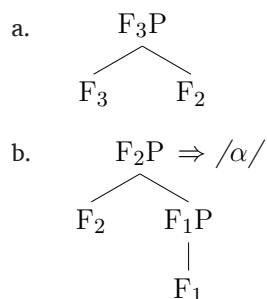
A completed subderivation is integrated into the main structure as a complex left branch and spelled out as a prefix with respect to F_2P:

---

20   Starke (2018) argues that spellout-driven movement always follows a specific algorithm. There are three basic steps in the algorithm, stay, spec-to-spec movement and roll-up movement.

21   Note that all the previous options have to be attempted first, before a subderivation can be spawned.

(56)

$$F_3P$$

$$/\beta/ \Leftarrow F_3P \qquad F_2P \Rightarrow /\alpha/$$

$$F_3 \quad F_2 \qquad F_2 \quad F_1P$$

$$| $$

$$F_1$$
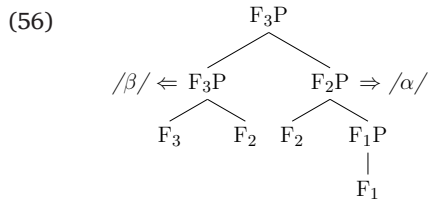
The spellout-driven movement and subderivation mechanisms are of great significance to the nanosyntactic theory, since they alter tree geometry and facilitate matching between syntactic structure and lexically stored trees. Moreover, these mechanisms reveal that there is a clear structural difference between prefixes and suffixes. Spellout-driven movement will lead to the formation of a suffix, which is a remnant constituent with a unary foot, while subderivation will create a prefix, which is a complex left branch with a binary foot:

(57)   Suffix vs. prefix
    a.   suffix

$$\text{Stem} \quad F_xP$$

$$|$$

$$F_x$$

    b.   prefix

$$F_yP \qquad \text{Stem}$$
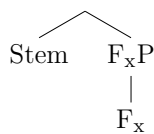
$$F_y \quad F_x$$
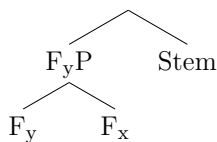
4.3 Nanosyntax – summary

As shown in this section, Nanosyntax introduces a new perspective on concepts such as the lexicon, the architecture of syntax and spellout. According to the nanosyntactic model of grammar, syntactic derivations arrange features according to the order specified by a cross-linguistically universal sequence (the fseq). Constituents formed by features are cyclically spelled out with lexical information provided by entries contained in the lexicon. An entry can spell out a piece of structure it matches or is overspecified for (the Superset Principle). In cases where two or more lexical entries are eligible to lexicalize a particular set of terminals, the one with the fewest superfluous features wins the competition (the Elsewhere Principle). If there is no lexical entry that can be used to spell out a structure, the syntactic system will employ syntactic transformations to obtain a lexicalizable configuration of features or spawn a subderivation that can be spelled out and integrated into the main structure as a complex left branch.[22]

The nanosyntactic model of derivation successfully explains syncretism as a phenomenon which stems from the basic rules of lexicalization, that is cyclic phrasal spell-out and the Superset Principle. Syncretism arises whenever two or more phrasal nodes can be spelled out with the same lexical entry, as permitted under the Superset Principle. Furthermore, the nanosyntactic spellout system accounts for the *ABA generalization, as the ABA pattern is not possible under the Elsewhere Principle. As shown in example (48), the Elsewhere Principle guarantees that the spellout system will always choose the most specific lexical entry for each cycle, which rules out the possibility of a lexical entry matching non-adjacent phrasal nodes.

Lastly, the spellout-driven movement and subderivation mechanisms reveal how suffixes and prefixes are formed. Syntactic transformations that are applied when a piece of syntactic structure does not match any lexical entry in the lexicon result in the formation of suffixes and prefixes. Spellout-driven movement will create a suffix with a unary foot, while subderivation will form a prefix with a binary foot.
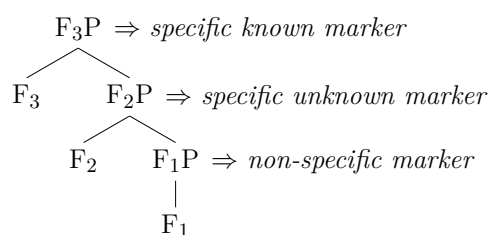
---

22   According to Starke (2018), the spellout algorithm is as follows: stay (and spell out), move – spec-to-spec, move – roll-up, backtrack and lastly subderive.

In the next section, I will explain in detail the derivation of all the patterns of syncretism attested in the studied language sample (AAA, ABC, AAB and ABB) using the analytical tools of Nanosyntax outlined in this section. It also seems worth noting at this point that the nanosyntactic framework proves to be useful not only in the case of indefinite markers. So far, the approach has been successful in analyzing the phenomenon of syncretism in many other grammatical domains such as participles (Starke 2006), case (Caha 2009), spatial adpositions (Pantcheva 2011), negation markers (De Clercq 2013), demonstratives (Lander & Haegeman 2016), personal pronouns (Vanden Wyngaerd 2018), wh-pronouns (Wiland 2018; Vangsnes 2013), complementizers (Baunaz & Lander 2018a), verbal prefixes in Slavic (Wiland 2012; Tolskaya 2018), class prefixes in Bantu (Taraldsen 2010; Taraldsen et al. 2018) and internal structure of verbs (Taraldsen-Medova & Wiland 2019; Wiland 2019).

5 Analysis

The methodological tools provided by the nanosyntactic framework, i.e. cyclic phrasal spell-out regulated by the Superset and Elsewhere Principles, allow us to propose a clear and coherent model of derivations for the syntactic structures corresponding to non-specific, specific unknown and specific known indefinite markers. As shown in the previous sections, a comparison of data from 45 languages reveals that the three types of markers are derived on the basis of a universal structural sequence. The attested patterns of syncretism and the *ABA generalization indicate that this sequence is a hierarchy based on structural containment. Syntactically, this kind of hierarchy can be represented as a structure consisting of three elements, where F_1, F_2 and F_3 show the levels of syntactic embedding:[23]

(58)   Indefinite hierarchy

$$F_3P \Rightarrow \textit{specific known marker}$$
$$F_3 \quad F_2P \Rightarrow \textit{specific unknown marker}$$
$$F_2 \quad F_1P \Rightarrow \textit{non-specific marker}$$
$$F_1$$

The layers of syntactic structure forming the non-specific, specific unknown and specific known indefinite markers have to be assembled in a particular consecutive order (dictated by the fseq). F_2 will always be preceded by the merger of F_1, and F_3 will be merged only once F_1 and F_2 have been assembled. In consequence, since spell-out targets only phrasal nodes, the three types of indefinite markers will constitute the lexical items corresponding to different subconstituents of the hierarchy.

The proposed model of the syntactic structure underlying non-specific, specific unknown and specific known indefinite markers can be used to explain all the attested patterns of syncretism and the absence of the ABA pattern. The factor that determines the pattern for each language is the number of lexical entries that can match the indefinite structure. Below, I use English, Yakut and Latin to illustrate the derivation of the attested patterns of syncretism (AAA, ABB and AAB), and Russian to explain cases with no syncretism. Additionally, the analysis shows the steps necessary to spell out indefinite markers as either prefixes or suffixes. The analysis of Russian shows how a prefix can be derived from a suffix, while Latin reveals the derivation of a suffix from a prefix.
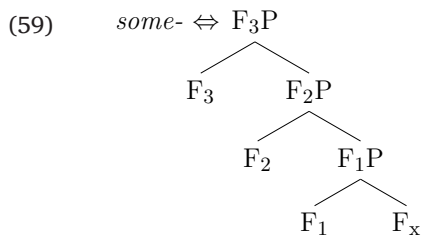
5.1 English (AAA pattern)

As seen in *Table 2*, English has only one lexical exponent for all three of the indefinite markers, namely (*some-*):

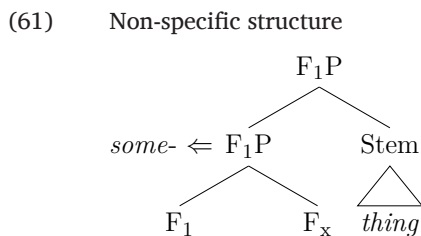|	non-specific	specific unknown	specific known	pattern
English	some-	some-	some-	AAA

Table 2 English.

This means that in English, there is only one lexical entry that is used to lexicalize the sequence $[F_1, F_2, F_3]$ and all its subsets. In other words, under the Superset Principle, the entry matching the whole hierarchy will also spell out $[F_1, F_2]$ and $[F_1]$:
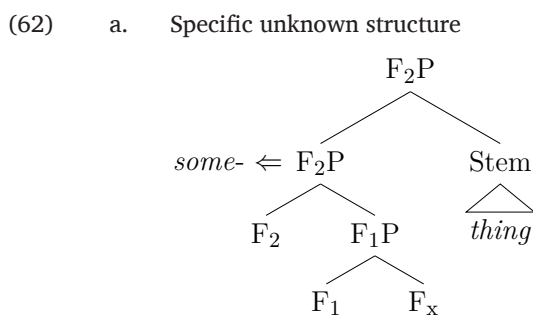
(59)   *some-* $\Leftrightarrow$ F_3P

$$
\begin{array}{c}
some\text{-} \Leftrightarrow F_3P \\
\diagup\diagdown \\
F_3 \quad F_2P \\
\diagup\diagdown \\
F_2 \quad F_1P \\
\diagup\diagdown \\
F_1 \quad F_x
\end{array}
$$

The three indefinite markers in English appear as prefixes with respect to the categorical stem, which means that their structure is formed through subderivation. When the first indefinite layer (F_1) is merged on top of the stem, it will not be spelled out due to the lack of a matching entry in the lexicon. The lexical entry in (59) will not match the derived structure:

(60)

$$
\begin{array}{c}
F_1P \Rightarrow ? \\
\diagup\diagdown \\
F_1 \quad Stem \\
\quad\triangle \\
\quad thing
\end{array}
$$
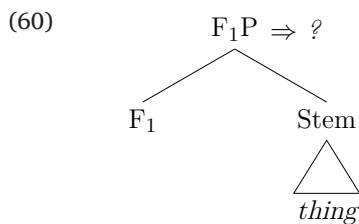
Subsequent attempts to lexicalize the structure through spellout-driven movement (spec-to-spec and roll-up) and then backtracking will also not provide a lexicalizable tree geometry, at which point the derivation will have to resort to subderivation. The merger of F_1 will be undone and a new parallel derivation will be created. This subderivation will begin with the merger of F_1 with a copy of the last succesfully merged feature from the main sequence to form the minimal derivation structure (Caha et al. 2019). The properties of this feature are not relevant to the analysis, which is why it will be labeled F_x. The subderivation will be spelled out with the lexical entry in (59) and integrated into the main structure in the following way:

(61)   Non-specific structure

$$
\begin{array}{c}
F_1P \\
\diagup\diagdown \\
some\text{-} \Leftarrow F_1P \quad Stem \\
\diagup\diagdown \quad \triangle \\
F_1 \quad F_x \quad thing
\end{array}
$$

With the subsequent indefinite layers (F_2 and F_3), the steps described above will be repeated.[24] After multiple failed attempts to spell out the next feature provided by the fseq (movement and backtracking), the derivation will eventually backtrack to the stem and spawn a derivation providing the required features.[25] Each time, the subderivation will be spelled out as a subset of the entry in (59):

(62)   a.   Specific unknown structure

$$
\begin{array}{c}
F_2P \\
\diagup\diagdown \\
some\text{-} \Leftarrow F_2P \quad Stem \\
\diagup\diagdown \quad \triangle \\
F_2 \quad F_1P \quad thing \\
\diagup\diagdown \\
F_1 \quad F_x
\end{array}
$$

---

24   The next feature will be merged on top of the structure derived in the previous cycle.

25   The feature that the system is trying to spell out in the current cycle and all the features that have been undone. The derivation has to follow the order provided by the fseq (see Section 4).

b. Specific known structure

$$F_3P$$

some- $\Leftarrow$ F_3P Stem

F_3 F_2P *thing*

F_2 F_1P

F_1 F_x
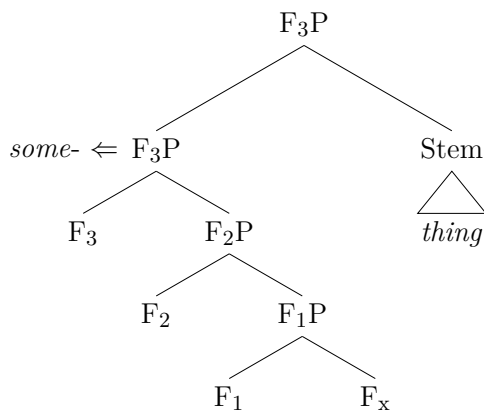
Starke (2018) suggests an alternative to the derivation of prefixes described above. According to this proposal, subderive may be considered such a costly operation that a parallel derivation should be kept active as long as possible, instead of being integrated into the main spine immediately after providing the feature that the system wants to spell out in the current cycle. This means that it may not be necessary to repeat the spellout algorithm (stay and spell out, move, backtrack and finally subderive) to derive each layer of the indefinite structure. After the first indefinite layer is subderived (F_1P), the subderivation will be extended and also provide F_2 and F_3 (depending on the intended indefinite structure). This will generate the same results as the derivation process show above.
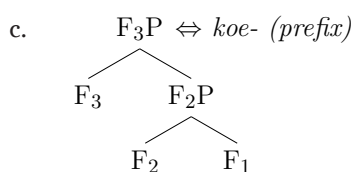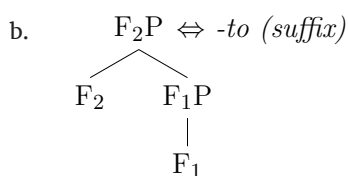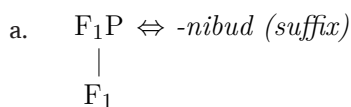
5.2 Russian (no syncretism)

The indefinite marker paradigm for Russian contains three separate forms. This is shown in *Table 3*:

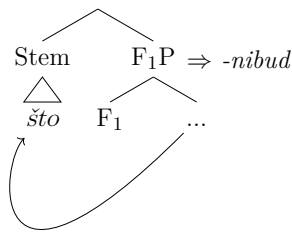|	non-specific	specific unknown	specific known	pattern
Russian	-nibud	-to	koe-	ABC

This means that there are three separate lexical entries which lexicalize the indefinite hierarchy. Each lexical entry matches a different subconstituent of the hierarchy:

(63) Lexical entries

a. $F_1P \Leftrightarrow$ *-nibud (suffix)*

F_1

b. $F_2P \Leftrightarrow$ *-to (suffix)*

F_2 F_1P

F_1

c. $F_3P \Leftrightarrow$ *koe- (prefix)*

F_3 F_2P

F_2 F_1

Since the first two indefinite markers (non-specific and specific unknown) are suffixes, they are spelled out through the displacement of the stem to a position above the indefinite projections, in line with the spellout-driven movement mechanism (see Section 4.2). The phonological exponent matched with the specific unknown structure will overwrite the exponent corresponding to the smaller non-specific structure:

(64)   a.   Non-specific structure

Stem $F_1P \Rightarrow$ *-nibud*

što F_1 ...

    b.   Specific unknown structure

Stem $F_2P \Rightarrow$ *-to*

*što* F_2

... F_1P

F_1 ...
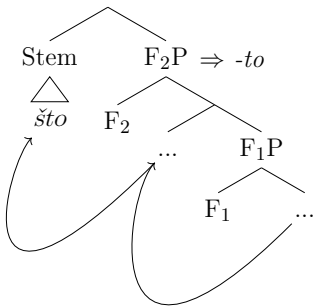
In contrast with the non-specific and specific unknown markers, the specific known marker *koe-* is a prefix, which means that its underlying structure should constitute a subderived phrase containing all three layers of the indefinite hierarchy. The derivation of this structure will begin with the merger of (F_3) on top of the tree shown in (64-b). The resulting structure will not match any entry in the lexicon:

(65) $F_3P \Rightarrow$ *?*

F_3

Stem $F_2P \Rightarrow$ *-to*

što F_2

... F_1P

F_1 ...
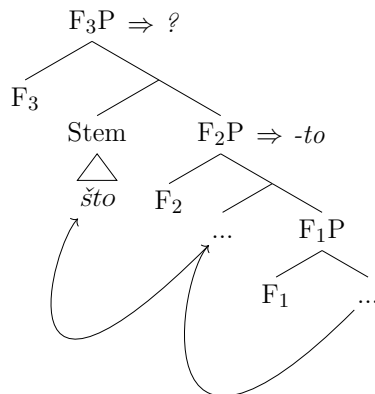
Since spellout-driven movement will also fail to produce a lexicalizable tree geometry, the spellout mechanism will force syntax to create a subderivation. However, a subderivation spawned after layers F_1 and F_2 have been assembled will not lead to the formation of the desired structure. To properly derive a left branch containing the sequence [F_1, F_2, F_3], the syntactic system has to resort to backtracking, undo the merger of all the previously derived indefinite layers (F_1 and F_2), and then create a subderivation. Layers F_1 and F_2 will be merged at the bottom of the subderivation to form the minimal binary structure and the parallel derivation will then remain active until the F_3 is provided. The resulting structure will be integrated into the main sequence as a complex left branch (a prefix):[26]

(66) F_3P

$F_3P \Rightarrow$ *koe-*   Stem

F_3 F_2P *što*

F_2 F_1

---
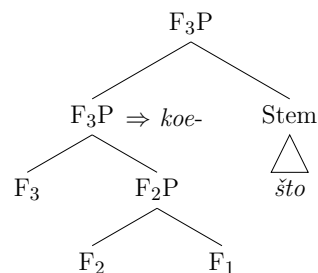
26   No feature from the main sequence will appear in the subderivation, since there is no need to derive a prefix containing only the first layer of the indefinite hierarchy (F_1).
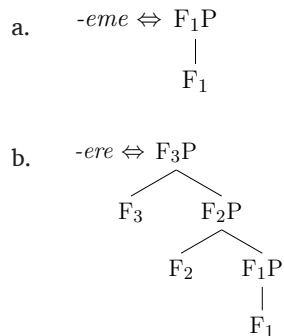
5.3 Yakut (ABB pattern)

Yakut represents a group of languages in which the specific unknown and specific known markers are syncretic to the exclusion of the non-specific marker. The forms are shown in *Table 4*:

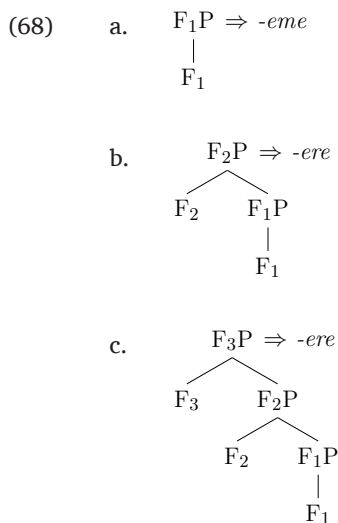|	non-specific	specific unknown	specific known	pattern
Yakut	-eme	-ere	-ere	ABB

The observed pattern (ABB) results from the fact that the indefinite hierarchy is lexicalized by the exponents of two lexical entries:

(67)　　Lexical entries

　　　a.　$-eme \Leftrightarrow F_1P$
　　　　　　　|
　　　　　　F_1

　　　b.　$-ere \Leftrightarrow F_3P$
　　　　　╱＼
　　　　F_3　　F_2P
　　　　　　　╱＼
　　　　　　F_2　　F_1P
　　　　　　　　　|
　　　　　　　　F_1

The first entry, i.e. (67-a), will spell out only the smallest subset of the hierarchy. This entry will always be matched with F_1P, since the entry in (67-b) contains superfluous features and will be ignored under the Elsewhere principle. When F_2P and F_3P are assembled, (67-b) becomes the only matching entry. This means that the indefinite structure will be spelled out the following way:

(68)　a.　$F_1P \Rightarrow -eme$
　　　　　　|
　　　　　F_1

　　　b.　$F_2P \Rightarrow -ere$
　　　　╱＼
　　　F_2　　F_1P
　　　　　　　|
　　　　　　F_1

　　　c.　$F_3P \Rightarrow -ere$
　　　　╱＼
　　　F_3　　F_2P
　　　　　　╱＼
　　　　　F_2　　F_1P
　　　　　　　　|
　　　　　　　F_1

Because all three indefinite markers in Yakut are suffixes, they will be spelled out as remnant constituents formed through the displacement (roll-up) of the categorical stem:

(69)
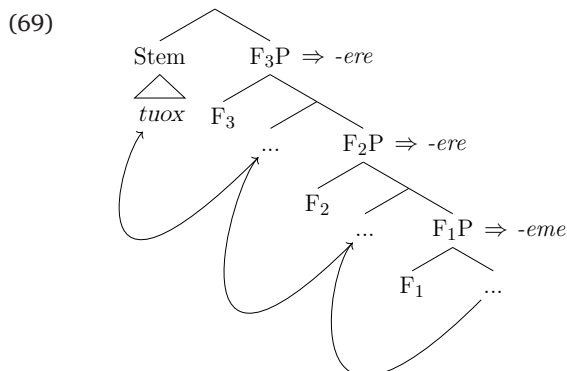
5.4 Latin (AAB pattern)

Latin represents the AAB pattern; the non-specific and specific unknown markers are syncretic to the exclusion of the specific known marker. *Table 5* shows the indefinite marker paradigm for Latin:

|	non-specific	specific unknown	specific known	pattern
Latin	ali-	ali-	-dam	AAB

The examples below show the two lexical entries that lexicalize the indefinite hierarchy in Latin:

(70)　　Lexical entries

　　a.　　$ali\text{-} \Leftrightarrow F_2P$

　　b.　　$\text{-}dam \Leftrightarrow F_3P$
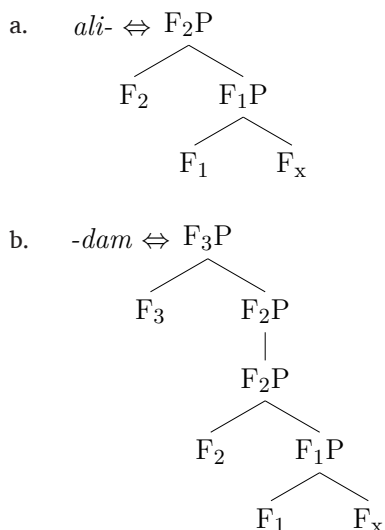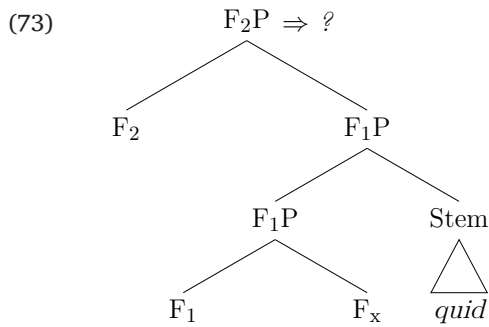

The non-specific layer of the indefinite hierarchy is subderived in Latin and lexically realized as a prefix (*ali-*). The subderivation procedure is applied, since all other attempts to spell out F_1, i.e. movement of the stem and movement after backtracking to the previous merge cycle, will not produce a lexicalizable tree geometry:
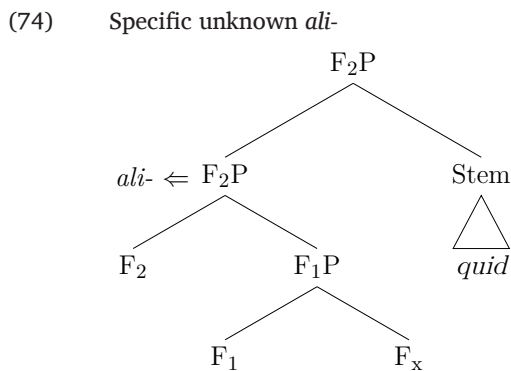
(71)　　$F_1P \Rightarrow$?
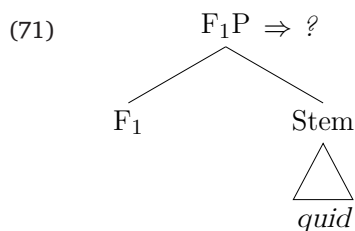
The only way to obtain a lexicalizable tree is through subderivation where F_1 will be merged with a copy of the last successfully lexicalized feature from the main spine (F_x) to form a minimal strucutre. The resulting constituent can be lexicalized as a subset of (70-a) (*ali-*):
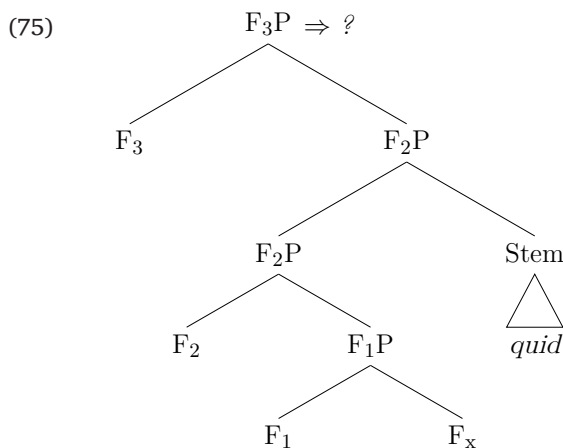
(72)　　Non-specific *ali-*
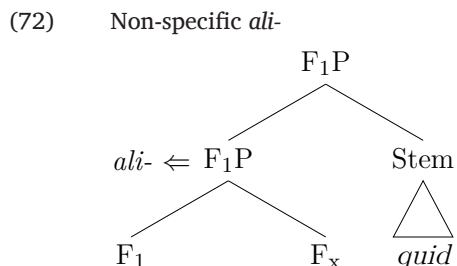
Next, to create the specific unknown structure, F_2 will be merged on top of F_1P. The resulting structure will however not match any of the available lexical entries. Any subsequent spellout operations (i.e. move and backtrack) will also not generate a tree that can be spelled out with (70-a) or (70-b):

(73)

$$F_2P \Rightarrow \, ?$$

                    F_2P ⇒ ?
                   /   \
                 F_2   F_1P
                           /   \
                        F_1P   Stem
                       /   \   △
                     F_1   F_x   *quid*

To create a lexicalizable structure containing F_2, it is again necessary to backtrack to the stem and spawn a subderivation.[27] The subderived constituent will be formed with F_1 and F_x as the foot (which can be spelled out as a subset of (70-a)) and grown to contain F_2.[28] The resulting constituent will be spelled out with (70-b)

(74)   Specific unknown *ali-*

                         F_2P
                        /   \
          ali- ⇐ F_2P   Stem
                  /   \   △
                F_2   F_1P   *quid*
                      /   \
                    F_1   F_x

The specific known marker *-dam* is a suffix, which means that it is derived through spellout-driven movement. It is not possible to spell out F_3 immediately after merge, which will trigger the spellout algorithm (spellout-driven movement):

(75)

                    F_3P ⇒ ?
                   /   \
                 F_3   F_2P
                           /   \
                        F_2P   Stem
                       /   \   △
                     F_2   F_1P *quid*
                           /   \
   F_1   F_x

The simplest way in which the syntactic system can obtain a tree geometry where all the layers of the indefinite hierarchy form a constituent is to extract the stem constituent (the complement of F_2P) from the phrase projected through the subderivation of the left branch F_2P (the prefix). Note that the proposed movement is not in line with the spellout algorithm proposed in Starke (2018), which predicts only spec-to-spec and roll-up movements before the spellout system

---

27   Unless we accept the proposal from Starke (2018), in which case, the subderivation spawned to spell out F_1 may remain active and provide F_2.

28   As mentioned in Section 4, features have to be merged in the order dictated by the fseq. This means that F_1 has to be merged before F_2.

has to resort to backtracking and finally subderivation. However, since the extraction of a previously lexicalized constituent is the most straightforward way of deriving a suffix (*-dam*) from a prefix (*-ali*), it is reasonable to suggest that constituent extraction may constitute the last movement option that may be applied (after spec-to-spec and roll-up movements) before backtracking has to be used.[29] The suffix (F_3P) is lexicalized with the phonological exponent of entry (70-b):[30]

(76)
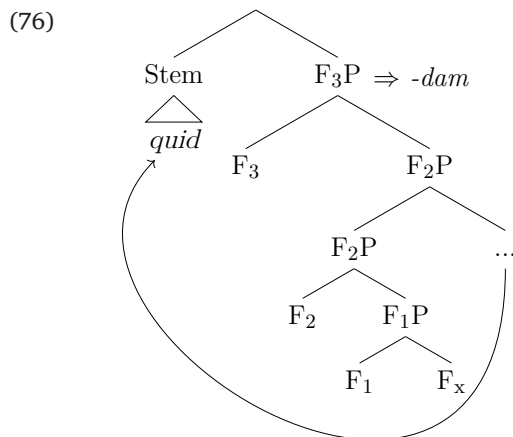
Note that despite being a suffix, the constituent formed through the extraction of the stem will have a binary foot. This is due to the fact that the previously derived left branch (F_2P) will remain as a remnant constituent inside the newly created suffix.[31]

5.5 Analysis – summary

The nanosyntactic model of grammar can be successfully used to create a coherent analysis of the derivation of indefinite markers. Non-specific, specific unknown and specific known indefinite markers, which appear as either prefixes of suffixes on categorial stems, constitute phonological exponents corresponding to subsets of a hierarchy of indefinite features. Language-specific restrictions on lexicalization (the contents of the lexicon) are responsible for the fact that languages spell out the indefinite hierarchy in different ways. However, despite this variation in lexicalization of indefinite markers, the nanosyntactic spellout system, regulated by the Superset and Elsewhere Principles, guarantees that the *ABA rule is not broken.

A matter that may be considered in terms of future research is the fact that indefinite markers also do not seem to violate *ABA when it comes to their morphological forms (prefixes/suffixes). In other words, there are no languages in the studied language sample in which the non-specific and specific known markers are suffixes (or prefixes) to the exclusion of the specific unknown marker. However, even if we are dealing with a genuine pattern in this case, at this point, it is not clear what its cause may be.[32]

6 Loose ends

The following section is devoted to the data that raise additional questions concerning the morphosyntax of indefinite markers. While not particularly relevant in the contexts of the presented analysis, some of the collected examples are still worth discussing with regard to the internal structure of indefinite markers and may potentially lead us to a number of interesting conclusions. Below, I address the issues of wh-indefinites and paradigm gaps.

---

29 Compare with Wiland (2019: 65–69), which suggests the possibility of extraction from a subconstituent.

30 The empty F_2P (after the stem has been extracted)node will not be ignored since it is projected by the F_2P constituent.

31 The AAB pattern appears to be quite rare (I have found only one language with this pattern, i.e. Latin). As noticed by an anonymous reviewer, Bobaljik (2012) characterizes this pattern as rare and difficult to derive (using the methodology of Distributed Morphology). The fact that the AAB pattern does not appear very frequently may be connected with the complexity of the syntactic structure necessary to derive it (a remnant prefix inside a suffix).

32 The possible existence of a prefix/suffix pattern was pointed out to me by an anonymous reviewer. I would like to thank them for this observation.
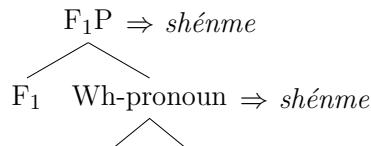
6.1 Indefinites syncretic with wh-pronouns

In a number of languages, we observe syncretism between interrogative pronouns and indefinite pronouns. Consider examples from Hopi, Dyirbal, Khmer, colloquial Dutch, colloquial German and Mandarin Chinese:[33]

(77)   Wh-word – indefinite pronoun syncretism
      a.   *hak* – 'who'/'somebody' (*Hopi*)
      b.   *minya* – 'what'/'something' (*Dyirbal*)
      c.   *naa* – 'where'/'somewhere' (*Khmer*)
      d.   *was* – 'what'/'something' (*colloquial German*)
      e.   *wat* – 'what'/'something' (*colloquial Dutch*)
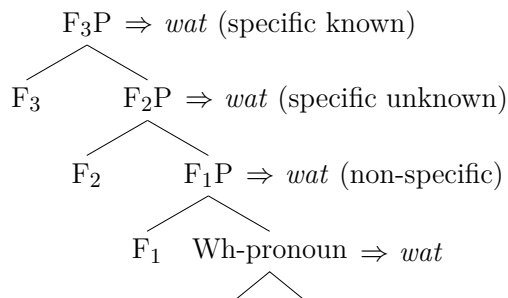   f.   *shénme* – 'what'/'something' (*Mandarin Chinese*)

Examples of this kind of syncretism can be analyzed as cases where interrogative pronouns are spelled out as syncretic subsets of indefinite pronouns. In other words, the same lexical entry can be used to spell out the interrogative structure as well as the interrogative structure with indefinite layers on top of it. Consider, the following example from Mandarin Chinese, where only non-specific pronouns are syncretic with interrogatives (cf. Li 1992 and Lin 1998).[34] The non-specific indefinite (*shénme*) will be lexicalized with the same lexical entry as the interrogative pronoun (*shénme*)

(78)   *Mandarin Chinese – shénme* 'what'/'something'

$$\text{F}_1\text{P} \Rightarrow \textit{sh\'enme}$$
$$\text{F}_1 \quad \text{Wh-pronoun} \Rightarrow \textit{sh\'enme}$$

In colloquial Dutch, pronouns for all three indefinite functions are syncretic with the interrogative pronouns. This means that for a particular category, the non-specific, specific unknown, specific known and interrogative pronouns are all spelled out as subsets of a single lexical entry:

(79)   *Dutch – wat* 'what'/'something'

$$\text{F}_3\text{P} \Rightarrow \textit{wat} \text{ (specific known)}$$
$$\text{F}_3 \quad \text{F}_2\text{P} \Rightarrow \textit{wat} \text{ (specific unknown)}$$
$$\text{F}_2 \quad \text{F}_1\text{P} \Rightarrow \textit{wat} \text{ (non-specific)}$$
$$\text{F}_1 \quad \text{Wh-pronoun} \Rightarrow \textit{wat}$$

6.2 Paradigm gaps

The studied data sample contains a number of languages in which we observe the absence of pronouns for one or more indefinite functions. The presence of such paradigm gaps makes these languages largely irrelevant in proving the proposed claim concerning the syntactic structure of indefinites. However, it should also be stressed that no language shown below can be used as an example against it. I discuss languages with paradigm gaps in an attempt to provide a possible explanation for this phenomenon:[35]

---

33   Examples taken from Dixon (1972: 265) – Dyirbal, Hengeveld et al. (2020) – Dutch, Huffman (1967: 153–6) – Khmer, Malotki (1979: 110) – Hopi and Haspelmath (1997: 170).

34   See also Section 6.2.

35   For a full description of the data shown in Table 6, see Appendix 2.

	non-specific	specific unknown	specific known	pattern
Kannada	-aadaruu	-oo	–	AB
Quechua	-pis/-pas	-chi/-cha	–	AB
Mandarin Chinese	wh-pronoun	–	–	A
Irish	–	–	–	–
Swahili	–	–	–	–
Filipino	–	–	–	–
unattested				ABA

Table 6 Paradigm gaps.

There is yet another reason why it may be worth taking a closer look at paradigm gaps. There is a possibility that gaps in indefinite pronoun paradigms form a certain pattern. The currently available data seem to indicate that if only one indefinite pronoun type is missing, it is the most complex one (specific known), and if two indefinite pronoun types are absent, it should be the specific ones. However, since my data concerning languages with paradigm gaps is limited, I am not able to confirm the existence of an actual pattern in this case.

6.3 Paradigm gaps – possible explanations

The data in *Table 6* can be accounted for in two ways. The first possibility that we may want to consider is that the indefinite hierarchy is smaller (reduced) in some languages. In other words, one or more layers of the indefinite hierarchy could be "cut off". However, the main issue with this approach is that it does not seem to be consistent with the available data. In Irish, for example, despite the lack of pronominal indefinites, the three indefinite functions can be expressed with the noun modifier *éigin* 'some'. Some other languages mentioned in this section, e.g. Swahili, Filipino and Chinese, also have indefinite modifiers corresponding to the indefinite functions that cannot be expressed pronominally. This shows that the indefinite hierarchy is unlikely to be absent or reduced in these languages. If so, then sequence reduction cannot be the correct explanation of the data.

A solution that takes into account indefinite noun modifiers is closely connected with the nanosyntactic lexicon. If paradigm gaps do not stem from variation in syntax, they may be caused by the lexicon.[36] In other words, a language does not have indefinite pronouns of a given type if it lacks a lexical entry (or entries) that is necessary to spell out the indefinite structure corresponding to that type. This, of course, means that features which cannot be spelled out as part of an indefinite pronoun may still appear inside other lexicalizable structures, such as indefinite modifiers. Additionally, this solution to the problem of paradigm gaps explains the regularity that can be seen in *Table 6*. If only one indefinite pronoun type is missing, it will be the specific known type. If two pronoun types are absent, it will be the specific types (specific known and unknown). This kind of pattern should be expected if paradigm gaps are caused by the unavailability of lexical entries. Under the nanosyntactic rules of lexicalization, a paradigm gap could only appear if a structure lacks a matching lexical entry and cannot be spelled out by a larger entry matching its superset. If a given structure can be spelled out, then all its subsets are also lexicalizable. For this reason we should never see a gap for an indefinite function if pronouns for a more complex function (or functions) are available.

7 Conclusion

The nanosyntactic framework of grammar allows us to create a coherent analysis of the internal structure and derivation of non-specific, specific unknowns and specific known indefinite markers. The presented model of the internal syntactic structure of these markers is based on a study of indefinite marker syncretism in a 45-language sample. *Table 7* and *Table 8* contain examples illustrating the discovered patterns:

---

36   Interestingly, this leads us back to the idea mentioned in Section 4, namely that linguistic variation does not stem from syntax, but is instead caused by differences on the level of the lexicon.
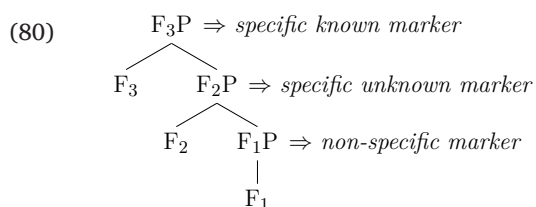
	non-specific	specific unknown	specific known	pattern
English	some-	some-	some-	AAA
Polish	-ś	-ś	-ś	AAA
Japanese	-ka	-ka	-ka	AAA
Korean	-nka	-nka	-nka	AAA
Lezgian	-jaťani	-jaťani	-jaťani	AAA
Romanian	-va	-va	-va	AAA
Bulgarian	nja-	nja-	nja-	AAA
Serbo-Croatian	ne-	ne-	ne-	AAA
Czech	ně-	ně-	ně-	AAA
Slovak	nie-	nie-	nie-	AAA
Maltese Czech	xi-	xi-	xi-	AAA
Hungarian	vala-	vala-	vala-	AAA
Hebrew	-šehu	-šehu	-šehu	AAA
Turkish	bir-	bir-	bir-	AAA
Latvian	kaut-	kaut-	kaut-	AAA
Yakut	-eme	-ere	-ere	ABB
Georgian	-me	-ɣac	-ɣac	ABB
Ossetic	is-	-dær	-dær	ABB
Latin	ali-	ali-	-dam	AAB
Russian	-nibud	-to	koe-	ABC
Lithuanian	-nors	kaž-	kai-	ABC
unattested				ABA

Table 7 Indefinite markers – syncretism.

	non-specific	specific unknown	specific known	pattern
Kannada	-aadaruu	-oo	–	AB
Chinese	wh-pronoun	–	–	A
Swahili	–	–	–	–
Irish	–	–	–	–
Filipino	–	–	–	–
unattested				ABA

Table 8 Paradigm gaps.

The attested patterns of syncretism lead us to the conclusion that the non-specific, specific unknown and specific known indefinite markers should correspond to a hierarchy of syntactic containment comprising at least three layers of structure. Each indefinite marker type is derived as a different subset of the hierarchy:

(80)

$F_3P \Rightarrow$ *specific known marker*

F_3 $F_2P \Rightarrow$ *specific unknown marker*

F_2 $F_1P \Rightarrow$ *non-specific marker*

F_1

The nanosyntactic principles of lexicalization account for all patterns of syncretism found in the studied language sample and the fact that the ABA pattern remains unattested. Syncretism between non-specific, specific unknown and specific known indefinite markers arises whenever a language has a lexical entry matching two or more contiguous layers in the hierarchy. As for the ABA pattern, this pattern should not be expected to appear due to the Elsewhere Principle, which guarantees that lexical entries cannot match non-contiguous phrasal nodes.

Abbreviations

ACC – accusative

CONV – converb

DAT – dative

DIR – direction

FUT – future

HORT – hortative

IMPERF – imperfect

IMPV – imperative

INDEF – indefinite

INF – infinitive

INSTR – instrumental

Neg – negation

NOM – nominative

PAST – past

PERF – perfect

PL – plural

POL – polite

PRES – present

PT – particle

Q – quastion

Additional files

The additional files for this article can be found as follows:

- **Appendix 1.** Examples of specific and non-specific indefinite pronouns from 45 languages. DOI: *https://doi.org/10.5334/gjgl.1233.s1*
- **Appendix 2.** A description of the data concerning Section 6.2 (Paradigm gaps) and a brief discussion of languages with multiple indefinite pronoun paradigms. DOI: *https://doi.org/10.5334/gjgl.1233.s2*

Competing interests

The author has no competing interests to declare.

Author Affiliations

Jakub Dekier *orcid.org/0000-0001-8478-9042*
Adam Mickiewicz University in Poznań, PL

References

Afanas'ev, Petr S. & Luka N. Xaritonov. 1968. *Russko-jakutskij slovar'*. Moscow: Sovetskaja Ènciklopedija.

Aloni, Maria & Angelika Port. 2013. Epistemic indefinites cross-linguistically. In Yelena Fainleib, Nicholas LaClara & Yangsook Park (eds.), *Proceedings of NELS* 41. 29–43. Amherst, MA: GLSA Publications.

Baunaz, Lena & Eric Lander. 2018a. Deconstructing categories syncretic with the nominal complementizer. *Glossa: a journal of general linguistics* 3(1). 31. DOI: *https://doi.org/10.5334/gjgl.349*

Baunaz, Lena & Eric Lander. 2018b. Nanosyntax: the basics. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring Nanosyntax*, 13–84. Oxford University Press. DOI: *https://doi.org/10.1093/oso/9780190876746.001.0001*

Baunaz Lena & Eric Lander. 2019. Ontological Categories. In *The Unpublished Manuscript*, 1–18. *https://ling.auf.net/lingbuzz/003993*.

Baunaz Lena, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.) 2018. *Exploring Nanosyntax*. Oxford Studies in Comparative Syntax. Oxford University Press. DOI: *https://doi.org/10.1093/oso/9780190876746.001.0001*

Bobaljik, Jonathan. 2007. *On Comparative Suppletion*. Ms., University of Connecticut.

Bobaljik, Jonathan. 2012. *Universals in Comparative Morphology: Suppletion, Superlatives, and the Structure of Words*. Cambridge, MA. MIT Press. DOI: *https://doi.org/10.7551/mitpress/9069.001.0001*

Caha, Pavel. 2020. *Nanosyntax: some key features*. *https://ling.auf.net/lingbuzz/004437*.

Caha, Pavel. 2009. *Nanosyntax of Case*. PhD thesis, University of Tromsø.

Caha, Pavel, Karen De Clercq & Guido Vanden Wyngaerd. 2019. The fine structure of the comparative. *Studia Linguisitca* 73(3). 470–521. DOI: *https://doi.org/10.1111/stul.12107*

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In Kenneth Hale & Samuel J. Keyser (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 1–52. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford University Press.

Cinque, Guglielmo & Luigi Rizzi. 2008. The Cartography of Syntactic Structures. *CISCL Working Papers on Language and Cognition* 2. 43–59. DOI: *https://doi.org/10.1093/oxfordhb/9780199544004.013.0003*

Croft, William. 1983. Quantifier scope ambiguity and definiteness. *Berkeley Linguistic Society* 9. 25–36. DOI: *https://doi.org/10.3765/bls.v9i0.2018*

De Clercq, Karen. 2013. *A unified syntax of negation*. Univ. of Ghent PhD diss.

Dixon, R. M. W. 1972. *The Dyirbal Language of North Queensland*. Cambridge, Cambridge University Press. DOI: *https://doi.org/10.1017/CBO9781139084987*

Eremina, Olga. 2012. *The Semantics of Russian Indefinite Pronouns: Scope, Domain Widening, Specificity, and Proportionality and Their Interaction*. PhD dissertation. Michigan State University.

Fábregas, Antonio. 2007. An Exhaustive Lexicalisation Account of Directional Complements. *Nordlyd* 34(2). 165–199. DOI: *https://doi.org/10.7557/12.110*

Haspelmath, Martin. 1997. *Indefinite Pronouns*. Oxford, Clarendon Press.

Hengeveld, Kees, Sabine Iatridou & Floris Roelofsen. 2020. *Quexistentials and Focus*. *https://ling.auf.net/lingbuzz/005601*.

Heringer, James T. 1969. Indefinite noun phrases and referential opacity. *Chicago Linguistic Society* 5. 89–97.

Huffman, Franklin E. 1967. *An outline of Cambodian grammar*. Ph.D. dissertation, Cornell University.

Karttunen, Lauri. 1976. Discourse referents. In James D. McCawley (ed.), *Notes from the linguistic underground* (Syntax and Semantics, 7), 363–85. New York: Academic Press. DOI: *https://doi.org/10.1163/9789004368859_021*

Kayne, Richard. 2005. *Movement and Silence*. Oxford Studies in Comparative Syntax. Oxford University Press. DOI: *https://doi.org/10.1093/acprof:oso/9780195179163.001.0001*

Kiparsky, Paul. 1973. 'Elsewhere' in phonology. In Stephen Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 93–106. Holt, Rinehart Winston, New York.

Kozhanov, Kirill. 2015. Lithuanian indefinite pronouns in contact. In Peter Arkadiev, Axel Holvoet & Björn Wiemer (eds.). *Contemporary Approaches to Baltic Linguistics*, 465–490. Mouton de Gruyter.

Kulaev, Nikolaj. 1958. *Mestoimenija v sovremennom literaturnom osetinskom jazyke*. Ordž onikidze: Severo-osetinskoe kniž noe izdatel'stvo.

Lander, Eric & Liliane Haegeman. 2016. The Nanosyntax of Spatial Deixis. *Studia Linguistica* 72(2). 362–427. DOI: *https://doi.org/10.1111/stul.12061*

Li, Yen-Hui Audrey. 1992. Indefinite wh in Mandarin Chinese. *Journal of East Asian Linguistics* 1(2). 125–55. DOI: *https://doi.org/10.1007/BF00130234*

Lin, Jo-Wang. 1998. On Existential Polarity WH-Phrases in Chinese. *Journal of East Asian Linguistics* 7(3). 219–255. DOI: *https://doi.org/10.1023/A:1008284513325*

Malotki, Ekkehart. 1979. *Hopi-Raum: Eine sprachwissenschaftliche Analyse der Raumvorstellungen in der Hopi-Sprache*. Tübingen, Niemeyer.

Onenko, Sulungu. 1986. *Nanajsko-russkij slovar'*. Moscow: Russkij jazyk.

Pantcheva, Marina. 2011. *Decomposing path: The nanosyntax of directional expressions*. Tromsø: University of Tromsø Phd dissertation. *http://ling.auf.net/lingbuzz/001351*.

Pilka, Alfonsas. 1984. *Neopredelennye determinativy litovskogo jazyka (vsopostavlenii s anglijskim)*. Candidate's dissertation, Vilnius State University, Vilnius.

Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In Liliane Haegeman (ed.), *Elements of grammar*, 281–337. Springer, Dordrecht. DOI: *https://doi.org/10.1007/978-94-011-5420-8_7*

Sharahsenidze, Nino. 2018. Teaching Georgian as a second language: indefinite pronouns and modality. In *Indian Journal of Medical Ethics*, 126–128. DOI: *https://doi.org/10.22333/ijme.2018.110019*

Starke, Michal. 2001. On the inexistence of specifiers and the nature of heads. In Adriana Belletti (ed.), *Structures and beyond* 3. 252–268. Oxford University Press.

Starke, Michal. 2006. *The nanosyntax of participles*. Lectures at the 13th EGG summer school, Olomouc.

Starke, Michal. 2009. Nanosyntax A short primer to a new approach to language. In Peter Svenonius, Gillian Ramchand, Michal Starke & Knut Tarald Taraldsen (eds.), *Nordlyd* 36(1). 1–6. Tromsø: CASTL. DOI: *https://doi.org/10.7557/12.213*

Starke, Michal. 2011. Towards elegant parameters: Language variation reduces to the size of lexically stored trees. In Carme Picallo (ed.), *Linguistic Variation in the Minimalist Framework*. Oxford University Press.

Starke, Michal. 2014. Cleaning up the lexicon. *Linguistic Analysis*, 245–256.

Starke, Michal. 2018. Complex Left Branches, Spellout and Prefixes. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring Nanosyntax*, 349–363. Oxford University Press. DOI: *https://doi.org/10.1093/oso/9780190876746.003.0009*

Taraldsen, Knut T., Lucie Taraldsen Medová & David Langa. 2018. Class prefixes as Specifiers in Southern Bantu. *Natural Language Linguistic Theory* 36. 1339–1394. DOI: *https://doi.org/10.1007/s11049-017-9394-8*

Taraldsen, Tarald. 2010. The nanosyntax of Nguni noun class prefixes and concords. *Lingua* 120. 1522–1548. DOI: *https://doi.org/10.1016/j.lingua.2009.10.004*

Taraldsen-Medová, Lucie & Bartosz Wiland. 2019. Semelfactives are bigger than degree achievements. *Natural Language Linguistic Theory* 37(4). 1463–1513. DOI: *https://doi.org/10.1007/s11049-018-9434-z*

Tolskaya, Inna. 2018. Nanosyntax of Russian verbal prefixes. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring Nanosyntax.*, 205–236. New York: Oxford University Press. DOI: *https://doi.org/10.1093/oso/9780190876746.003.0008*

Ubrjatova, Elizaveta Ivanovna (ed.). 1982. *Grammatika sovremennogo jakutskogo literaturnogo jazyka*. Vol. 1. Moscow, Nauka.

Vanden Wyngaerd, Guido. 2018. The feature structure of pronouns: a probe into multidimensional paradigms. In Lena Baunaz, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.), *Exploring Nanosyntax*, 409–447. Oxford University Press.

Vangsnes, Øystein A. 2013. Syncretism and functional expansion in Germanic whexpressions. *Language Sciences* 36. 47–65. DOI: *https://doi.org/10.1016/j.langsci.2012.03.019*

Wiland, Bartosz. 2012. Prefix stacking, syncretism, and the syntactic hierarchy. In Mojmir Docekal & Marketa Zikova (eds.), *Slavic languages in formal grammar*, 307–324. Berlin/Frankfurt am Main: Peter Lang.

Wiland, Bartosz. 2018. A note on lexicalizing 'what' and 'who' in Russian and in Polish. *Poznan Studies in Contemporary Linguistics* 54(4). 573–604. DOI: *https://doi.org/10.1515/psicl-2018-0023*

Wiland, Bartosz. 2019. *The spell-out algorithm and lexicalization patterns: Slavic verbs and complementizers*. Berlin: Language Science Press.