Jakub Dekier

365356

Assertive indefinite pronouns: a morphosyntactic analysis

Twierdzące zaimki nieokreślone: analiza morfoskładniowa

Rozprawa doktorska napisana

na Wydziale Anglistyki

Uniwersytetu im. Adama Mickiewicza w Poznaniu

pod kierunkiem dr hab. Bartosza Wilanda

Poznań, 2022

Acknowledgements

Completing this project would not have been possible without the contribution and support of many people.

First, I would like to thank my supervisor Bartosz Wiland for his help during the past six years. Thank you for reading so many different versions of this dissertation and mercilessly pointing out my mistakes. I will not forget your advice.

I would also like to extend my gratitude to Michal Starke who agreed to appear as the expert for my doctoral presentation. I really appreciate the fact that you came all the way to Poznań and offered detailed suggestions.

Jacek Witkoś is another person who supported me during all these years as a Ph.D. student. Thank you, I was lucky to have such a great advisor.

Next, I would like to mention those who helped me with my research. Adobo, Elisavet, Daniela, Hagamablabla, Liu, Maria, Maria, Marie, Michaela, Opper-pax, Sako, Space, Swedebro, Šarka, Twittar, Qok, and many others. Thank you!

Lastly, I have to thank my friends and family. If it had not been for your support, I would not be where I am today. It was your presence that made it possible from me to deal with the many obstacles I had to overcome over the past few years.

OŚWIADCZENIE

Ja, niżej podpisany

Jakub Dekier

---

przedkładam rozprawę doktorską

pt. Assertive indefinite pronouns: a morphosyntactic analysis
Asertoryczne zaimki nieokreślone: analiza morfoskładniowa

na Uniwersytecie im. Adama Mickiewicza w Poznaniu

i oświadczam,
że napisałem/am ją samodzielnie.

Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, niekorzystałem/am z pomocy innych osób, a w szczególności nie zlecałem/am opracowania rozprawy lub jej istotnych części innym osobom, ani nie odpisywałem/am tej rozprawy lub jej istotnych części od innych osób. Jednocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

Poznań, 01.06.2022 Jakub Dekier

(miejscowość, data) (czytelny podpis)

Table of contents

List of tables

List of figures

List of abbreviations

ABL - ablative

ACC - accusative

ADJ - adjective

ALL - allative

ABS - absolutive

ART - article

BPG - best possible grounds

CAUS - causative

CL - classifier

COMIT - comitative

CON - conative

CONJ - conjectural

CONT - continuous

CONV - converb

DAT - dative

DECL - declarative

ELA - elative

ERG - ergative

FUT - future

GEN - genitive

HAB - habitual

HORT - hortative

IMPERF - imperfect

IMPV - imperative

INESS - inessive

INF - infinitive

INSTR - instrumental

LIM - limitative

LOC - locative

NEG - negation

NMLZ - nominalizer

NOM - nominative

PART - particle

PTCP - participle

PRTV - partitive

PAST - past

PERF - perfect

PFV - perfective

PL - plural

POL - polite

POSS - possessive

PREP - preposition

PRES - present

PROG - progressive

Q - question

REFL - reflective

REL - relative

SBJV - subjunctive

SG - singular

SS - same subject

TEMP - temporal

TOP - topic

VAL - validational

Introduction

In this dissertation, I analyze the syntactic composition of assertive indefinite pronouns and their corresponding markers. Some examples showing indefinite pronouns which belong in this category are given below:

(1) English

a. *some-thing*

b. *some-body*

c. *some-where*

(2) Polish

a. *co -ś*
what INDEF
'something'

b. *kto -ś*
who INDEF
'somebody'

c. *gdzie -ś*
where INDEF
'somewhere'

(3) Hebrew (Haspelmath 1997: 298)

a. *ma -šehu*
what INDEF
'something'

b. *mi -šehu*
who INDEF
'somebody'

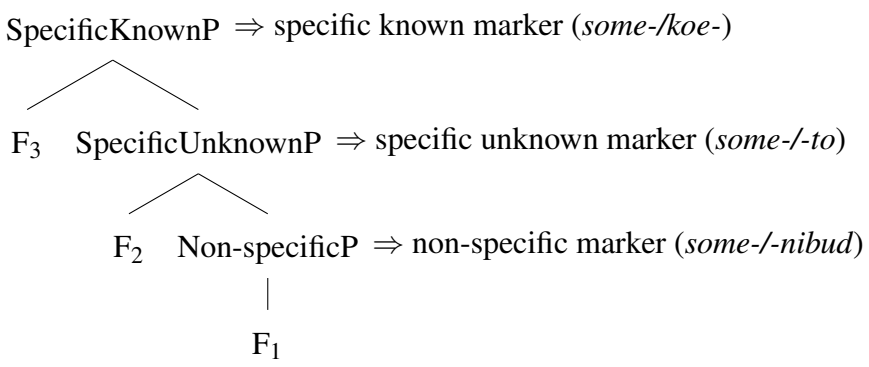
c. *eyfo -šehu*
where INDEF
'somewhere'

As argued in Haspelmath (1997) and confirmed by a cross-linguistic comparative study presented in this thesis, indefinite markers used to form assertive indefinite pronouns can be divided into three types, that is the non-specific, specific unknown and specific known indefinite markers. Each of the three types of markers expresses a different indefinite function, as illustrated by the following examples:

(4) a. *I want **some-thing** to eat. It doesn't matter what.* non-specific function
b. *There is **some-thing** in the box. I don't know what.* specific unknown function
c. *I have **some-thing** to tell you. Guess what!* specific known function

Subsequently, I report that syncretism may arise between the lexical forms of indefinite markers corresponding to the non-specific, specific unknown and specific known indefinite functions. The distribution of the attested forms of syncretism is limited to adjacent forms in a particular order of indefinite markers. Therefore, following the *ABA generalization of Bobaljik (2007, 2012), which states that syncretism targets only contiguous items in an ordered set of forms, we can propose a paradigm for the non-specific, specific unknown and specific known indefinite markers. The order of the indefinite markers in the paradigm is argued to reflect the internal syntactic composition of these forms:

(5) Indefinite hierarchy



As suggested by the distributions of syncretic forms, markers corresponding to the non-specific, specific unknown and specific known indefinite functions realize three different sets of features based on a cumulative hierarchy of syntactic features (5). The proposed representation of the internal structure of these markers is also supported by their functional composition since specificity is the prerequisite for knowledge of the speaker. In other words, the specific unknown indefinite marker introduces the property of specific reference, whereas specific unknown marker expresses both specificity and knowledge of the speaker (a uniquely identifiable referent). The non-specific marker is thus considered to be the simplest type.

While the structure in (5) is taken to be cross-linguistically universal, it can be phonologically realized in different ways. The attested forms of syncretism result from different language-specific lexicalizations of the indefinite hierarchy. For example, the hierarchy may be represented by only one phonological exponent, as in English, or as many as three separate forms, as in Russian.

The presented analysis of non-specific, specific unknown and specific known indefinite markers draws from the methodology of nanosyntax (Starke 2009) and its core tenets such as the Superset Principle, the Elsewhere Principle and phrasal spellout. With the use of these methodological tools, it can be illustrated how the cross-linguistically attested forms of syncretism stem from the number of lexical entries in the form of lexically stored syntactic trees which match the indefinite hierarchy in a given language. When a lexically stored tree matches a particular structure assembled in syntax, the phonological exponent corresponding to that tree will lexicalize that structure. Syncretism may arise since the Superset Principle permits a single lexically stored tree to match a particular structure and its syntactic subsets. If multiple lexical entries compete for insertion, the most specific entry will be chosen, as dictated by the Else-

where Principle. In consequence of this system, the number of phonological exponents that realize the non-specific, specific unknown and specific known indefinite functions will reflect the number of lexically stored trees that can match the indefinite hierarchy.

Furthermore, the application of the nanosyntactic system of lexicalization reveals how indefinite markers are generated as prefixes or suffixes, and how structural growth may result in the presence of both prefixes and suffixes in the indefinite pronoun paradigm of a given language. Examples showing a change in the morphological position of the indefinite marker with the change of its type are provided in (6) and (7):

(6) Lithuanian indefinite prefix and suffix (Haspelmath 1997: 275)

a. *kas* ***-nors***
what INDEF
'something' (non-specific)

b. ***kaž-*** *kas*
INDEF what
'something' (specific unknown)

(7) Latin indefinite prefix and suffix (Haspelmath 1997: 254)

a. ***ali-*** *quid*
INDEF what
'something' (specific unknown)

b. *quid* ***-dam***
what INDEF
'something' (specific known)

The dissertation is structured the following way. In Chapter 1, I discuss indefinite pronouns as a grammatical category and their traditionally recognized subtypes. This chapter also presents the general morpho-syntactic and semantic characteristics of indefinite pronouns that allow us to recognize and define them as a category. Lastly, I introduce a typology of assertive indefinite pronouns based on the analysis presented in Haspelmath (1997).

Chapter 2 is devoted to the typology of non-specific, specific unknown and specific known indefinite pronouns introduced in Haspelmath (1997). On the basis of examples from

multiple languages, I explain the semantic and syntactic properties of these three types of indefinite pronouns. This typological analysis also includes a discussion of the distribution of non-specific, specific unknown and specific known indefinite pronouns across different semantic and syntactic contexts.

Chapter 3 presents the data collected for the purpose of the analysis. I show the forms of syncretism that may arise cross-linguistically in the paradigm of the non-specific, specific unknown and specific known indefinite marker types. In the second part of the chapter, I discuss examples of languages which do not have pronominal forms corresponding to one or more of the three types of indefinite markers/functions.

In Chapter 4, I discuss the methodology which is used to propose an analysis of the internal structure of the non-specific, specific unknown and specific known indefinite markers. The chapter begins with information concerning the core tenets of the nanosyntactic model of grammar such as the one feature - one head maxim and phrasal spellout of syntactic structures. Subsequently, I introduce the idea of the nanosyntactic lexicon and illustrate the core methodological concepts in nanosyntax such as the Superset Principle and the Elsewhere Principle. The last part of the chapter is devoted to mechanisms of the nanosyntactic derivation system such as spellout-driven movement, backtracking and subderivation.

Chapter 5 contains the main analysis which shows the emergence of the attested patterns of syncretism through the lexicalization of the proposed indefinite hierarchy. I begin the chapter with a brief discussion of the three types of indefinite pronoun bases and the idea that they express basic concepts known as ontological categories. The main part of the chapter is devoted to the examples illustrating the derivation of different forms of indefinite marker syncretism and the formation of indefinite suffixes and prefixes. Additionally, it is shown how the nanosyntactic system of lexicalization can deal with portmanteau indefinites. I will argue that the non-specific, specific unknown and specific known indefinite markers are formed in a cumulative way (feature stacking). This means that the specific unknown marker is formed by adding a feature to the non-specific marker, and the specific known marker by adding an additional feature to the specific unknown marker. At the same time, I will argue against a non-cumulative representation in which the three markers are not syntactically contained within one another.

Chapter 6, which is the last main chapter of the dissertation, discusses a number of additional ideas that were not explored in the main analysis. This chapter briefly presents issues such as unproductive indefinite markers, paradigm gaps and overlapping indefinite pronoun series.

The thesis concludes with a summary which reiterates the main points of the analysis. I also address to possibility of extending the proposed analysis of assertive indefinite markers to learn more about the internal structure of other types of indefinites such as NPI and free choice pronouns.

Chapter 1: Indefinite pronouns

1.1. Indefinite pronouns: the basics

Expressions belonging to the grammatical category known as indefinite pronouns (later also indefinites for short) can be found in many languages of the world. The sentences below show a few examples of indefinite pronouns and their use in English, Polish and Hungarian:

(8) English

a. ***Someone*** *called. I didn't know who it was.*

b. *Bring me* ***something*** *to eat, please.*

c. *She probably went* ***somewhere****. She'll be back in a minute.*

d. *You should visit us* ***sometime****.*

e. *Fortunately, he* ***somehow*** *managed to pass the exam.*

(9) Polish[1]

a. ***Ktoś*** *dzwonił. Nie wiem kto.*
someone called not know.1SG who
'**Someone** called. I don't know who.'

b. *Przynieś mi* ***coś*** *do jedzenia.*
bring.IMPV me something for eating
'Bring me **something** to eat.'

c. *Powinieneś nas* ***kiedyś*** *odwiedzić.*
should.2SG us sometime visit
'You should visit us **sometime**.'

[1] All Polish examples are presented according to my native knowledge of the language, JD.

d. ***Jakoś*** *zdał* *egzamin*.
somehow passed.3SG exam
'**Somehow** he passed the exam.'

(10) Hungarian[2]

a. *Van* ***valamim*** *a számodra. Találd ki, hogy mi az*!
there.is something the for.you guess that what that
'I have **something** for you. Guess what!'

b. *Mary hozzá akar menni* ***valakihez*** *az Egyesült Államokból*.
Mary wants go-INF someone-ALL the united states-ELA
'Mary wants to marry **someone** from the United States.'

c. *Hozz nekem* ***valamit*** *enni! Teljesen mindegy, hogy mit*.
bring-IMPV me something eat-INF completely doesn't.matter that what
'Bring me **something** to eat. It doesn't really matter what.'

From a general perspective, indefinite pronouns, as a category, comprise a variety of pronominal forms that denote indefinite reference. This description indicates that there are two main characteristics that allow us to identify indefinite pronouns among other grammatical categories, namely indefiniteness and pronominality. I will discuss these two properties below.

The idea of indefiniteness appears in the literature as a notion based on general intuition, speakers' judgments and the pragmatics of indefinite expressions, rather than one particular definition (cf. Abbott 2008, Haspelmath 1997: 12-14, Hawkins 2015). The concept of indefiniteness is notoriously difficult to explain, and there is no straightforward formulation that can fully capture what it means to be indefinite. For example, a referent that represents or exemplifies a larger group (it is one of many) is considered indefinite (11-a), but so is a referent that does not refer to any particular member of a group (11-b):

(11) a. ***A teacher*** *entered the classroom.* (there are many teachers in the school and one of them entered the classroom)

b. *Sometimes,* ***a teacher*** *comes and inspects the locker room.* (it may be a different teacher each time)

[2]If not indicated otherwise, all linguistic data have been obtained from native speakers.

I discuss the issues concerning the indefiniteness of indefinite pronouns in more detail in Section 2.1.. For now, however, I will use the following simple diagnostic to show that the intuition we have about indefinite pronouns is not misguided. Indefinites of the kind shown in example (8) resemble nominal phrases with indefinite articles. Such phrases are considered unambiguously indefinite:

(12) a. ***A person*** *came*. ⇒ ***Someone*** *came*.
b. *There is* ***a thing*** *on the table*. ⇒ *There is* ***something*** *on the table*.
c. *Let's go* ***to a place***. ⇒ *Let's go* ***somewhere***.

Although it cannot be said that the sentences in each pair are semantically identical, either of them could be uttered in a context where the referent is not considered definite (a particular identifiable entity) (cf. Abbott 2008). In contrast, expressions with definite articles may not be replaced by indefinite pronouns:

(13) a. ***The man*** *came*. ⇒ *****Someone*** *came*.
b. *He put* ***the thing*** *on the table*. ⇒ *He put ****something*** *on the table*.
c. *She visited* ***the place*** *where they sell boats*. ⇒ *She visited ****somewhere*** *where they sell boats*.

Apart from expressing a form of indefiniteness, indefinite pronouns are, as the name suggests, pronominal, which means that they constitute a special category of lexical items that replace (indefinite) nominal expressions, for example:

(14) a. *a/some person* ⇒ *somebody*
b. *a/some thing* ⇒ *something*
c. *a/some place* ⇒ *somewhere*

However, note that indefinite pronouns do not replace only nominal phrases. In fact, they may replace quite a wide variety of expressions, which suggests that rather than pro-nouns they should be considered a diverse category of pro-forms (cf. Haspelmath 1997: 10-13). In the following sentences, *somewhere*, *somehow* and *sometime* will not replace nominal phrases:

(15) a. *Let's go **somewhere**. ⇒ Let's go **to some place**.*

b. *He **somehow** managed to pass the test. ⇒ He managed to pass the test **in some way**.*

c. *Visit us **sometime**. ⇒ Visit us **at some time**.*

In Polish, indefinites may replace an even larger variety of expressions than in English. The following indefinite pro-forms are absent from English:

(16) Polish

a. *Weź **któregoś** z tych.*
take.IMPV which.ACC-INDEF from these
'Take **(some) one** of these.'

b. *Musiał **którędyś** wejść.*
had.to.3SG which.way-INDEF enter
'He had to enter through **some way**.'

c. ***Skądś** go znam.*
from.where-INDEF him know.1SG
'I know him **from somewhere**.'

d. *Za drzwiami stoi **jakiś** człowiek.*
behind doors stands what.kind.of-INDEF man
'**Some** man is standing behind the door.'

The differences between indefinites and lexical expressions are revealed through their morphological, phonological and syntactic properties. The characteristic of indefinite pronouns that is easy to notice is that they generally consist of two main parts, an indefinite marker (in bold) and a base:

(17) Indefinite markers and bases

a. ***some**-thing*

b. *co-**ś*** (Polish, 'something')

c. *nani-**ka*** (Japanese, 'something')

d. *ima-**pis*** (Quechua, 'something')

e. *vala-**mi*** (Hungarian, ‘something’)

An indefinite marker is a functional part of an indefinite pronoun which conveys its indefinite meaning:

(18) English *some-* marker

a. ***some**-thing*
b. ***some**-one*
c. ***some**-where*

(19) Polish *-ś* marker

a. *co-**ś*** ‘something’
b. *kto-**ś*** ‘someone’
c. *gdzie-**ś*** ‘somewhere’

(20) Japanese *-ka* marker

a. *nani-**ka*** ‘something’
b. *dare-**ka*** ‘someone’
c. *doko-**ka*** ‘somewhere’

A single indefinite marker may be used to form multiple indefinite pronouns (a series), each with a different base:

(21) English *some-* series

a. *some-**body***, *some-**one***
b. *some-**thing***
c. *some-**where***, *some-**place***
d. *some-**how***
e. *some-**time***

(22) English *any-* series

a. *any-**body***, *any-**one***

b. *any-**thing***

c. *any-**where***

d. *any-**how***

e. *any-**time***

(23) Polish: interrogative-pronoun-based *-ś* series

a. ***kto**-ś* ‘somebody’

b. ***co**-ś* ‘something’

c. ***gdzie**-ś* ‘somewhere’

d. ***jako**-ś* ‘somehow’

e. ***kiedy**-ś* ‘sometime’

(24) Polish: interrogative-pronoun-based free choice *-kolwiek* series

a. ***kto**-kolwiek* ‘anybody’

b. ***co**-kolwiek* ‘anything’

c. ***gdzie**-kolwiek* ‘anywhere’

d. ***jak**-kolwiek* ‘anyhow’

e. ***kiedy**-kolwiek* ‘anytime’

Apart from an indefinite marker, every indefinite pronoun contains a base. In contrast to the indefinite marker, this part of an indefinite pronoun requires a more detailed explanation.[3] I will use the term indefinite pronoun base (later also base) to refer to forms that function as derivation stems for indefinite markers. For example, the bases for the marker *some-* in indefinites such as *something* an *somebody* are *thing* and *body*, respectively. Also note that indefinite pronoun bases may take different forms such as generic nouns, wh-pronouns or generic pronouns derived from the numeral *one*:

[3] I discuss indefinite pronoun bases in greater detail in Section 5.2.

(25) English: indefinite pronoun bases

a. *some-**thing*** (noun base)

b. *some-**where*** (wh-pronoun base)

c. *some-**one*** (pronominal *one* base)

The role of the base is to specify the category of the referent for the indefinite marker to modify. As referent categories, we should understand concepts that can be expressed with nouns such as *person*, *thing*, *place* or *manner* (cf. Baunaz and Lander 2018a, Haspelmath 1997: 21-31, Jackendoff 1983: 48-56). In an indefinite pronoun series, each referent category will be represented by a different base:

(26) Referent categories

a. person: *some-**body***/*some-**one***

b. thing: *some-**thing***

c. place: *some-**where***

d. manner: *some-**how***

e. time: *some-**time***

The number and form of available bases may vary from language to language. In Polish, for example, we observe types of bases which are not found in English (compare with (16)):[4]

(27) Polish

a. ***ile**-ś* ('some number/amount', ***ile*** 'how much/many')

b. ***jaki**-ś* ('some kind of', ***jaki*** 'what/what kind of')

c. ***który**-ś* ('one from a set', ***który*** 'which')

d. ***skąd**-ś* ('from somewhere', ***skąd*** 'from where')

Indefinite markers and their bases appear in two main configurations (cf. Haspelmath 1997: 27-32). An indefinite marker may be either a prefix (as in (25)) or a suffix (as in (27)) with respect to its base. However, it should be noted that there are indefinite pronouns that do not conform to the general form of indefinite pronouns (i.e. indefinite marker + base). In such cases, it is difficult or impossible to clearly distinguish the indefinite marker and the base.

[4]Indefinite pronouns in Polish are derived from interrogative pronouns.

Indefinites of this kind may be blends, portmanteaus or forms syncretic with other lexical items, for example interrogative pronouns and indefinite modifiers. I will call such forms special indefinites (Haspelmath 1997: 249, 284):

(28) Special indefinites

a. *alguien* 'somebody' (Spanish, a blend of *alguno* 'some' and *quien* 'who')

b. *kuch* 'something' (Hindi, portmanteau)

c. *shénme* 'something' (Chinese, shares its form with the interrogative pronoun *shénme* 'what')

d. *någon* 'somebody' (Swedish, from **ne-hwait-ik-hwarir* 'I don't know who', shares its form with *någon* 'some')

Another characteristic of indefinite pronouns that may distinguish them from lexical expressions is stress placement. In English, for example, indefinite pronouns will always receive initial stress, while in nominal phrases with indefinite modifiers the noun is stressed:

(29) Indefinite pronouns: stress placement

a. *some* ˈ*thing* → ˈ*something*

b. *some* ˈ*body* → ˈ*somebody*

c. *some* ˈ*time* → ˈ*sometime*

Lastly, indefinite pronouns may differ from nominal expressions with indefinite modifiers with respect to the placement of adjectives. As show in the examples below, adjectives in English, Polish and Spanish follow indefinite pronouns instead of preceding them, as in nominal expressions:

(30) English: indefinite pronouns with adjectives

a. *a/some* ***good*** *thing* → *something* ***good***

b. *a/some* ***nice*** *person* → *somebody* ***nice***

(31) Polish: indefinite pronouns with adjectives

a. ***dobra*** *rzecz ('a good thing')* → *coś* ***dobrego*** *('something good')*

b. ***miła*** *osoba ('a nice person')* → *ktoś* ***miły*** *('somebody nice')*

(32) Spanish: indefinite pronouns with adjectives

a. *una **buena** cosa* ('a good thing') → *algo **bueno*** ('something good')

b. *una **linda** persona* ('a nice person') → *alguien **lindo*** ('somebody nice')

While the discussed characteristics allow us to identify different kinds of indefinite pronouns, it should however be noted that examples provided so far show mainly indefinites of one particular type. Traditionally, indefinite pronouns such as English *someone* and *something* are considered to constitute the category of assertive (existential) indefinite pronouns due to the fact that they appear to assert the existence of an indefinite referent (Quirk et al. 1985: 376–377).[5] This kind of indefinites are often considered to be only one subtype within the category of indefinite pronouns. There are potentially three other kinds of indefinites that can be distinguished, namely NPI/affective pronouns (33), free choice pronouns (34) and negative pronouns (35) (Haspelmath 1997, Quirk et al. 1985: 376–377):

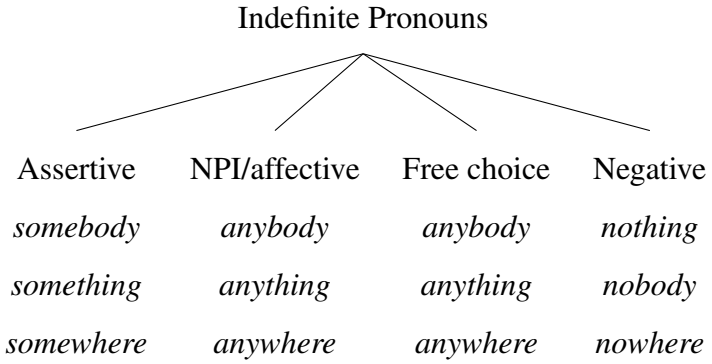


Figure 1. Indefinite pronouns: typology.

Affective/NPI indefinites are excluded from affirmative episodic sentences and require particular licensing environments, for example questions, sentences with sentential negators or nonveridical statements (cf. Giannakidou 2001, Giannakidou 2011a, Giannakidou 2011b, Haspelmath 1997: 33-37, Klima 1964):

(33) English: affective/NPI indefinites

a. *I don't think **anybody** saw me.*

b. *Is there **anything** we can do?*

c. **I saw **anybody** in the room.*

[5]The existential meaning of indefinite pronouns is discussed in Section 2.2.

d. **He put **anything** into the box.*

Pronouns expressing the free choice function are licensed in contexts which allow free selection from a number of alternatives, such as nonepisodic and nonveridical clauses (Giannakidou 2001, see also Giannakidou 2011a, Giannakidou 2011b, Haspelmath 1997: 48-51). They will be rendered ungrammatical in sentences where alternatives are not possible:[6]

(34) English: free choice indefinites

a. ***Anybody** can ride a bike.*

b. *You may take **whatever** you want.*

c. **Anybody** received an award.*

d. **I am going **anywhere**.*

Lastly, negative indefinite pronouns (also known as n-words) express negation and negate the entity they refer to (cf. Giannakidou and Zeijlstra 2017, Haspelmath 1997: 31-33):

(35) English: negative indefinties

a. ***Nobody** knows the answer.*

b. *He said **nothing**.*

c. ***Nowhere** else have I seen the whole story put down with such elegance and gusto.*

In some languages, for example in Polish (36), negative indefinites behave in a manner similar to NPIs in the sense that they have to be licensed by the presence of a sentential negator:[7]

(36) Polish: negative concord

a. ***Nikt** mnie **nie** zna.*
nobody me not knows
Nobody knows me.

b. **Nikt** mnie zna.*
nobody me knows
Intended: **Nobody** knows me.

[6]Note that in English free choice pronouns have the same form as NPI/affective pronouns.
[7]This is known as negative concord.

In contrast, negative indefinites in languages such as English may never appear with a sentential negator in the same clause:

(37) English

a. ***Nobody*** *knows me.*

b. ****Nobody*** *does* ***not*** *know me.*

It is a matter of debate whether the pronominal forms shown in (33), (34) and (35) should all be considered indefinite. Negative pronouns, for example, could be argued to express non-existence rather than indefinite reference. Similar doubts concerning indefiniteness can be raised with respect to free choice items, since they exhibit properties similar to those of universal pronouns such as *everything* and *everybody*.[8]

1.2. Assertive indefinite pronouns: the claim

In this thesis, I will focus on a grammatical category known as assertive indefinite pronouns (cf. Quirk et al. 1985: 376–377). As will be argued, this category of indefinites should be analyzed as comprising three separate types, namely:

- non-specific indefinite pronouns
- specific unknown indefinite pronouns
- specific known indefinite pronouns

[8]Quite interesting is also the structural similarity between indefinites and universal pronouns such as *every-thing* and *every-one* Universal pronouns may follow the same structural pattern as indefinites, that is a base stem combined with a functional marker:

(i) a. *some-thing*
b. *every-thing*

(ii) Hungarian (Rounds 2001: 138)

a. *minden- ki*
every who
'everybody'
b. *minden- hol*
every where
'everywhere'

This suggests that indefinite and universal pronouns together belong to an even broader category of expressions.

The proposed typology is based on the analysis of indefinite pronoun functions proposed in Haspelmath (1997), as well as on my own cross-linguistic study of the lexical forms of indefinite pronouns conducted for the purposes of this thesis. In languages such as English, the three discussed types of indefinite pronouns are all formed on the basis of a single indefinite marker *-some*. However, cross-linguistically, the non-specific, specific unknown and specific known indefinite functions may also be expressed with two or even three separate series of indefinite pronouns, each with its own indefinite marker. Compare how the three indefinite functions are expressed in English and Russian:

(38) English

a. ***some**-thing* (non-specific)

b. ***some**-thing* (specific unknown)

c. ***some**-thing* (specific known)

(39) Russian (Haspelmath 1997: 273)

a. *što* ***-nibud***
what INDEF
'something' (non-specific)

b. *što* ***-to***
what INDEF
'something' (specific unknown)

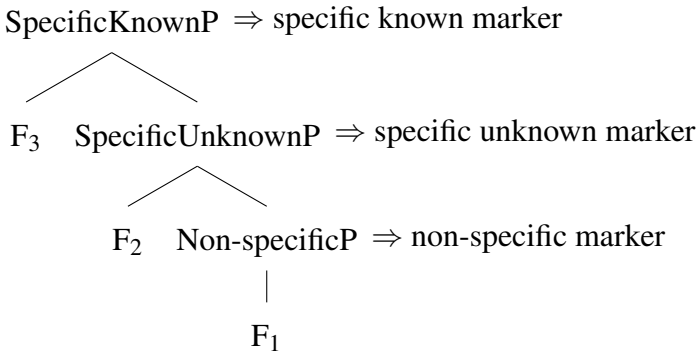
c. ***koe-*** *što*
INDEF what
'something' (specific known)

Taking into account these findings, I argue that cases in which a single indefinite marker (phonological exponent) is used to form indefinite pronouns corresponding to more than one indefinite function should be interpreted as instances of indefinite marker syncretism. Here, I understand the notion of syncretism as phonological identity of two or more non-identical syntactic structures (cf. Caha 2009: 6).

The phenomenon of syncretism can be understood and explained with the use of methodological tools of a theory of grammar known as nanosyntax (see Chapter 4.). On the basis of

the nanosytntactic approach to the lexicalization of syntactic structures, I argue that syncretic forms attested in indefinite marker paradigms cross-linguistically stem from the realization of a single hierarchy of syntactic features (an idea originally discussed in Dekier 2021). In the proposed hierarchy, the features F_1, F_2 and F_3 represent the levels of syntactic embedding:[9]

(40) Indefinite hierarchy



The non-specific, specific unknown and specific known indefinite marker types realize three different sets of features from the proposed hierarchy. Since the hierarchy is formed through syntactic feature cumulation (feature stacking), the non-specific, specific unknown and specific known indefinite markers arise in a particular order as the structure grows and are syntactically contained within one another. Thus, the non-specific marker realizes only the first layer of the hierarchy (F_1), while the specific unknown and the specific known markers correspond to the first two layers (F_1, F_2) and the whole hierarchy (F_1, F_2, F_3), respectively. We observe syncretism between different types of indefinite markers when a single phonological exponent is used to lexicalize more than just one layer of the indefinite hierarchy. An example of this is English (38), where only one form *some-* phonologically represents the non-specific, specific unknown and specific known indefinite functions. In contrast, we see no syncretism in Russian since the indefinite hierarchy is lexicalized with three different phonological exponents (39).

[9]In this analysis, I do not discuss the semantic content of the particular features in the hierarchy. The proposed representation is used to show the syntactic relation between the non-specific, specific unknown and specific known indefinite pronouns.

Chapter 2: Indefinite pronouns: typology

2.1. Three types of assertive indefinite markers

The following section discusses the typology of indefinite pronouns and their corresponding indefinite markers.[10] Out of the four types of indefinites mentioned in the previous section, in this thesis, I will focus solely on assertive indefinite pronouns (e.g. English *someone* and *something*). It should however be noted that despite their name, indefinites of this kind may appear in contexts other than just statements (declarative sentences). For example, English *some-* indefinites may appear in questions, requests, imperative clauses, or even with negation:

(41) English

a. *Do you want **something**?*

b. *Would you like **something** to drink?*

c. *John should go **somewhere** tomorrow.*

d. *It is important that **something** be done.*

e. *Give me **something** to read!*

f. *I didn't do **something**, and now I'm in trouble.*

For this reason, I will treat assertive indefinite pronouns as an umbrella term for indefinites other than NPI, free choice and negative pronouns. As will be shown, assertive indefinite pronouns, as a category, consist of three functionally distinct types of forms.

As indicated by a cross-linguistic comparison of indefinite pronouns presented in Haspelmath (1997), indefinites traditionally classified as assertive can be divided into three separate types based on their functional properties.[11] These three types are non-specific, specific un-

[10] For the time being, I will not discuss the second main part of an indefinite pronoun, that is the base. Examples of indefinite pronoun bases are *thing* in *some-thing*, *body* in *some-body* and *where* in *some-where*. Matters concerning the structure and properties of indefinite pronoun bases will be addressed in Section 5.2.

[11] I discuss Haspelmath's analysis in greater detail in Section 3.7.

known and specific known indefinite pronouns. Consider some simple examples that show the use of indefinite pronouns expressing the non-specific, specific unknown and specific known indefinite functions in English:

(42) Non-specific, specific unknown and specific known indefinite pronouns

a. *Bring me **something*** (non-specific) *to eat. I'm not very picky.*

b. *I heard he bought something* (specific unknown) *for his girlfriend. I'm not sure what it was.*

c. *I have **something*** (specific known) *for you. Try to guess what it is.*

Indefinite pronouns used in the non-specific function do not have a particular referent, as in (42-a). In contrast, indefinites expressing the specific unknown and specific known functions always refer to a particular entity (of the category specified by the base). The difference between specific unknown and specific known indefinite pronouns lies in the speaker's knowledge of the referent. Specific unknown indefinites have referents whose identity is not known to the speaker (42-b), while specific known indefinite pronouns refer to entities that the speaker can uniquely identify but does not want to speak openly about (42-c) (Haspelmath 1997: 37-47). Evidence for this kind of typology lies not only in the fact that indefinite pronouns such as the English *some-* series may receive three different interpretations, but comes primarily from languages in which each of the three indefinite functions is expressed by a different set of indefinites. Russian, for example, has a separate indefinite pronoun series for each of the three indefinite functions, the non-specific *-nibud*, specific unknown *-to* and specific known *koe-* series:

(43) Russian (Eremina 2012: 7, 72, 106)

a. *Za stenoj vsegda **kto-nibud** smejalsia.*
behind wall always someone laugh-PAST
'**Someone** (non-specific) was always laughing behind the wall (it might have been a different person each time).'

b. *Masha prigotovit **što-to** vkusnoje na uzhin.*
Masha cook-FUT something delicious for dinner
'Masha will cook **something** (specific unknown) delicious for dinner (and the speaker does not know what).'

c. *Ja nashla* ***koe-chto*** *interesnoje v etoj knige.*
I found something interesting in this book
'I saw **something** (specific known) interesting in this book (I know exactly what it was).'

In the sections below, I characterize in greater detail the functional and semantic properties of indefinite pronouns corresponding to the non-specific, specific unknown and specific known indefinite functions. The differences between these three types of indefinite pronouns will be the basis for the analysis presented in Chapter 5.

2.2. Specific unknown indefinite pronouns

Indefinite pronouns used in the specific unknown function refer to a particular entity of the kind specified by the base whose exact identity is however not known to the speaker (Eremina 2012: 37-70, Haspelmath 1997: 37-45).[12] Consider the following examples:

(44) English: specific unknown indefinite pronouns

a. ***Someone*** *is at the door. I don't know who they are.*
b. *She put* ***something*** *into the box. Maybe it was a comb.*
c. *He bought his hat* ***somewhere*** *in London. I don't remember the name of that place.*
d. ***Someone*** *was in my study. The cabinets are open.*
e. *She thinks that* ***something*** *fell into the well yesterday.*

In all of the cases above, the speaker has some particular referent in mind. The existence of that referent may be real or presupposed. For example, in (44-a), the speaker knows that there is a person at the door, while in (44-d), the speaker concludes on the basis of the available evidence that there is a person who entered the study. However, in all of the examples in (44) the exact identity of the referent always remains unknown to the speaker. In other words, the speaker lacks knowledge that would allow them to describe the referent in a definite way. The following examples also show how specific unknown indefinites are used when the speaker refers to a particular entity which they do not know, recognize or remember:

[12] For other works discussing the notion of specificity see Croft 1983, Enç 1991, Lavric 1990.

(45) a. ***Someone*** *called, but I don't know who.*

b. *He left his keys* ***somewhere*** *in the house, but he does not remember where.*

c. *There is* ***something*** *on the roof, but I can't see it clearly.*

d. *He* ***somehow*** *passed the test, but nobody knows how it was possible.*

e. *She said that the meeting would take place* ***sometime*** *next month, but she didn't know the exact date.*

Because specific unknown indefinites always have a particular referent, they naturally evoke existential interpretations. After all, the referent either actually exists or is assumed to exist within the frame established by the discourse. This can be shown through the following paraphrases with the existential *there*:

(46) a. ***Someone*** *is at the door.*

b. *There is* ***a person*** *at the door.*

(47) a. *She put* ***something*** *into the box.*

b. *There is* ***a thing*** *that she put into the box.*

(48) a. *Someone was in my study.*

b. *There is* ***a person*** *that was in my study.*

(49) a. *She thinks that* ***something*** *fell into the well yesterday.*

b. *She thinks that there is* ***a thing*** *that fell into the well yesterday.*

(50) a. *He believes that* ***someone*** *is under his bed.*

b. *He believes that there is* ***a person*** *under his bed.*

Indefinite pronouns expressing the specific unknown indefinite function commonly appear in declarative sentences such as the ones in (44). They are unlikely to be used in questions, except cases where the speaker presumes that the interlocutor wants to discuss a particular referent, or when a positive answer is expected (cf. Eremina 42-45):

(51) a. *Do you want* ***something***? (specific unknown) (A friend of yours has approached you and they seem to have a matter to discuss).

b. *Did you find **something***? (specific unknown) (A friend of yours has been looking for a gift for their mother. They are holding in item.).

Compare English with Russian, where the specific unknown marker *-to* may also appear in questions only in contexts in which specificity of the referent is presupposed (in such cases, the non-specific marker *-nibud* will be considered incorrect). The sentence below creates a presupposition that something bad actually happened, which is why the specific unknown marker *-to* is used:

(52) Russian (Eremina 2012: 43)

Kogda ja sdelal ***chto-to/*što-nibud*** *ne tak*?
when I do-PAST something not so
'When did I do **something** wrong?'

Specific unknown indefinite pronouns are also disallowed in imperative contexts since such clauses describe unrealized events. Imperative sentences with specific indefinites would also be considered illogical by interlocutors since the speaker cannot demand something that is a specific entity and at the same time withhold information that would allow the listener to identify an individual referent.[13] In consequence, the interlocutor is unable to fulfill the speakers request:[14]

(53) a. *Bring me* ****something*** (specific unknown) *that I need right now.*
b. *Tell* ****someone*** (specific unknown) *that the package is ready.*

Lastly, specific unknown indefinites are rarely used in sentences with sentential negation, and when they do appear in such sentences they may not fall under the scope of the negator. Eremina (2012: 40) gives the following example from Russian:
Context: "My friend complains that his family thinks he is a liar. I asked him why. And he says they are exaggerating. It only happened once, a long time ago, when he failed to tell something important to someone, and since then they don't trust him anymore even though no one really

[13] See Grice's cooperative principle (Grice 1975).

[14] Only non-specific indefinite pronouns are generally allowed in imperative sentences since they do not refer to a particular entity. The interlocutor does not need to know the exact identity of the referent if the speaker has mentioned (or demands) an unspecified referent of a given category (see Section 2.4.

remembers who that person was and what wasn't said" (Eremina 2012: 40).

(54) Russian (Eremina 2012: 40)

Ja ***komu-to što-to*** *ne skazal*.
I someone something not say-PAST
'I did not say *something* (specific unknown) to *someone* (specific unknown).'

2.3. Specific known indefinite pronouns

Indefinites of the specific known subcategory have referents that are not only specific but also known to the speaker. This means that the speaker can uniquely identify the referent but for some reason does not want to mention its/their exact identity. The knowledge of the listener/interlocutor is irrelevant (Haspelmath 1997: 45-51, cf. Eremina 2012: 105-112, Geist 2008: 159, Warfel 1972). The following examples show the use of indefinite pronouns in the specific known function in English:

(55) English: specific known indefinite pronouns

a. *John has made* ***something*** (specific known) *delicious for dinner* (I know what it is because I have seen it).
b. *I have* ***something*** (specific known) *for you. Try to guess what it is.*
c. *I met* ***someone*** (specific known) *on my way here* (I don't want to reveal who they were).
d. *I saw this book* ***somewhere*** (specific known) *around here, but I won't tell you where.*

Specific known indefinite pronouns are sometimes called *weak definites* (Paducheva 1985 as cited in Eremina 2012: 8) due to the fact that their referents are always a definite entity in the mind of the speaker. The referent is indefinite only in the context of the discourse (i.e. unknown to the listener). This causes specific know indefinites to be quite limited in their distribution, for example they are generally incompatible with imperative sentences and most questions:

(56) a. *Do you have* ****something*** (specific-known) *in that bag?*

b. *Could you bring me* ****someone*** (specific-known) *who can open this door?*

c. *Bring me* ****something*** (specific-known) *that I can put into the box.*

d. *Send* ****someone*** (specific-known) *who can memorize long messages.*

Specific known indefinite pronouns may appear in a questions, but usually only if the speaker is requesting assistance with a certain matter. In the example below, the speaker wants someone to do something for them, but they have not revealed the details yet. Naturally, the listener will ask for more information:

(57) A: *Can you do* ***something*** (specific known) *for me?*

B: *Yes, what is it?*

Imperative clauses with specific known indefinites are also possible if they are meant to be a challenge for the listener. The speaker knows what the requested items is, but the listener has to guess:
Context: You are playing a game with your friends. There are many colorful objects in the room. You notice a large red ball under the table. You say:

(58) *Bring me* ***something*** (specific known) *that is big and red.*

With negation, specific known indefinites behave like specific unknown forms, which means that they may never be interpreted within the scope of the negator. In the following example, the speaker discusses the reason why they were fired but does not want to reveal the details:

(59) *Why did they fire me? I didn't do* ***something*** (specific known) *important. I don't want to talk about it.*

The difference between known and unknown specific indefinite pronouns can be well illustrated by the following sets of sentences. Specific unknown indefinites always denote uncertainty or ignorance on the part of the speaker with respect to the identity of the referent, while specific known indefinites are limited to contexts where the speaker knows the exact identity of the referent:

(60) a. *He put **something*** (specific unknown) *into the box.*

- I have no idea what it is.

*- *I won't tell you what it is.*

b. *There is **someone*** (specific known) *behind the door.*

*- *I have no idea who it is.*

- I won't tell you who it is.

c. *He wants to go **somewhere*** (specific unknown) *tomorrow*.

- I don't known where.

*- *Guess where?*

d. *I want to go **somewhere*** (specific known) *tomorrow*.

*- *I don't know where.*

- Guess where?

2.4. Non-specific indefinite pronouns

Non-specific indefinite pronouns can be characterized as non-referential in the sense that they do not have particular referents (cf. Eremina 2012: 10-18, Haspelmath 1997: 37-45, Geist 2008: 159). In other words, the referent is not a concrete individual entity (either real or presupposed). The following examples show the use of non-specific indefinites in English:

(61) English: non-specific indefinite pronouns

a. *Bring me **something** to eat.* (I don't particularly care what.)

b. ***Someone** will probably come tomorrow to fix the tap.* (I'm not sure if anyone will actually come.)

c. *On Mondays, **someone** comes and takes away the trash.* (It is a different person every time.)

d. ***Something** may happen if you push the button.*

e. *She wants to go **somewhere**, but she still hasn't decided where.*

f. *He will probably buy **something** tomorrow.* (I'm not certain.)

g. *If you see **something*** (it doesn't matter what)*, tell me right away.*

In all the examples above, the referent is not identified as a specific member of the specified category (a particular person, thing or place, etc.). The speaker is unable or does not want

to narrow down the identity of the referent to a particular individual entity. The referent is uncertain, unknown, a random entity or does not exist at all in some cases. Generally, non-specific indefinites appear to refer only to a category variable (person, thing, place etc.), rather than a particular instance of that category. In consequence, the referent becomes an unspecified member of a chosen category.

The lack of a specific referent is the main characteristic that differentiates non-specific indefinite pronouns from the previously discussed types, namely specific indefinites (known and unknown). The difference between these two types can be illustrated by the following set of sentences (cf. Haspelmath 1997: 37-38):

(62) a. *Sally wants to marry **someone*** (specific) *from the US. She met him during a trip to the UK.*

b. *Sally wants to marry **someone*** (non-specific) *from the US because she is American herself.*

In example (62-a), Sally wants to marry a particular American person whose identity the speaker does not know. In contrast, in the second sentence (62-b), Sally does not want to marry a specific person from the US. It does not matter if a person that Sally would be willing to marry actually exists or not. The only thing we know is that if Sally were to marry someone, she would choose a person who comes from the US. Now consider a similar pair of sentences in Russian. Note that, unlike English, Russian marks the difference between specific and non-specific indefinites morphologically. The specific unknown function is expressed by the marker *-to* and the non-specific function by the marker *-nibud*:

(63) Russian: non-specific vs. specific indefinites (cf. Eremina 2012: 6-9)

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 the first sentence, only the non-specific pronoun is allowed since the described action happened

repeatedly and each time a different item was bought. In consequence, the indefinite pronoun lacks a specific referent (a particular singular entity) and the non-specific marker *-nibud* has to be used. The second case is different. This time the action occurred once and only one particular item was bought. Therefore, since the referent of the indefinite pronoun is a specific entity, the specific unknown marker *-to* has to be used.

Due to the fact that non-specific indefinites are non-referential, they are incompatible with discourse referents. This means that they cannot be antecedents for personal pronouns in present indicative clauses (cf. Karttunen 1976: 366):

(64) a. *Susan wants to buy **something*** (specific) *from that store. Unfortunately, it is expensive.*

b. *Susan wants to buy **something*** (non-specific) *nice for her sister (she still hasn't decided what to buy). *Unfortunately, **it** is expensive.*

c. *Mark wants to meet **someone*** (specific unknown) *from Cambodia. I've heard she is famous.*

d. *Mark wants to meet **someone*** (non-specific) *from Cambodia. *I've heard she is famous.*

It is also not possible to use non-specific indefinite pronouns in existential contexts. Paraphrases with existential constructions are impossible as well (cf. Heringer 1969: 90):

(65) a. *She put **something*** (specific) *into the box.*
*- There is **a thing** that she put into the box.*

b. *She wants to marry **someone*** (specific) *from the US.*
*- There is **a person** from the US whom she wants to marry.*

c. H*e often visits this store and buys **something*** (non-specific).
*- *There is **a thing** that he buys.*

d. *She wants to marry **someone*** (non-specific) *from the US because she is American.*
*- *There is **a person** from the US that she wants to marry.*

The fact that existential paraphrases are impossible with non-specific indefinite pronouns indicates that, unlike specific indefinites, they should be considered non-existential. The existence of a referent is not asserted by non-specific pronouns, which is a logical consequence of their

non-referentiality. In other words, obtaining the existential interpretation requires a specific referent that may exist. Of course, there may be sentences with non-specific indefinites for which the existence of potential referents will naturally be assumed. The referent is however never narrowed down to a particular entity. In (66-a), a referent is assumed to exist, while in (66-b), the existence of a referent is only a possibility:

(66) a. ***Somebody*** (non-specific) *comes and takes away the trash. It is a different person each time* (assumption: there are some people who take away the trash).
b. *I want* ***something*** (non-specific) *to cut this board with* (a possible referent, i.e. a tool for cutting boards, may exist or not).

Lastly, only specific indefinites may be paraphrased as nouns with modifiers such as *certain* which always indicates a specific referent. Indefinite pronouns used in the non-specific function may not be replaced by such phrases (Haspelmath 1997: 38, cf. Enç 1991):

(67) a. *She has put* ***something*** (specific) *into the box.* ⇒ *She has put* ***a certain thing*** *into the box.*
b. *He saw* ***someone*** (specific) *on the bridge.* ⇒ *He saw* ***a certain person*** *on the bridge.*
c. *He would like to eat* ***something*** (non-specific). *It doesn't matter what.* ⇒ *He would like to eat* ****a certain thing***.
d. *She meets* ***someone*** (non-specific) *at the office every day. She meets a lot of different people.* ⇒ *She meets* ****a certain person*** *at the office every day.*

Generally, non-specific indefinite pronouns require a particular grammatical context to be properly licensed, for example imperative clauses or certain irrealis structures (Haspelmath 1997: 39-45, cf. Croft 1983, Giannakidou 2011b). Such contexts do not indicate that the speaker has identified a concrete referent:

(68) a. *I wish I had* ***something*** (*specific/non-specific) *to write on. Unfortunately, I don't.*
b. *If only I knew* ***someone*** (*specific/non-specific) *from France. I would be able to translate this document.*

c. *Bring me **something*** (*specific/non-specific) *made of gold! Pick whatever you want.*

d. *Go **somewhere*** (*specific/non-specific) *where meals are less expensive!*

Non-specific indefinites are also typical in questions.[15] Note however that some languages have special NPI/affective indefinite forms that generally have to be used in interrogative contexts. The following example shows NPI/affective indefinites in English:

(69) NPI/affective pronouns

a. *Did you see **anyone**?*

b. *Is **anyone** here?*

c. *Have you heard **anything**?*

d. *Did you find **anything**?*

In contrast, languages such as Polish and Russian do not have separate NPI/affective indefinite series and use non-specific indefinites in questions:

(70) Russian (Eremina 2012: 43, 44)

a. *Ty nashel **što-nibud'** interesnoje?*
you find-PAST something interesting
'Did you find **anything** (something non-specific) interesting?'

b. *Ty kupila **što-nibud'** na uzhin?*
You buy-PAST something for dinner
'Did you buy **anything** (something non-specific) for dinner?'

(71) Polish

a. *Widziałeś **coś** w tej szafie?*
saw.2SG something in this closet
'Did you see **anything** (something non-specific) in this closet?'

[15] As mentioned in Section 2.2, specific indefinites may appear in questions only in very particular contexts.

b. *Czy **ktoś** go słyszał?*
Q someone him heard.3SG
'Did **anyone** (someone non-specific) hear him?'

In a similar way, conditional sentences are also a natural environment for non-specific indefinites:

(72) a. *If **someone** (non-specific) comes, let me know.*
b. *If **someone** (non-specific) gave me a lot of money, I would buy a ship.*
c. *If you find **something** (non-specific), give me a call.*

(73) Polish
a. *Jeśli **ktoś** przyjdzie, zadzwoń.*
if someone come-FUT.3SG call-IMPV.2SG
'If **someone** (non-specific) comes, call me.'
b. *Jeśli **coś** znajdziesz, niczego nie dotykaj.*
if something find-FUT.2SG nothing not touch-IMPV.2SG
'If you find **something** (non-specific), don't touch anything.'

Non-specific indefinites are usually disallowed in contexts which naturally impose the specific interpretation due to their logic, for example in past and ongoing present declarative sentences. Non-specific pronouns are possible if the sentence contains a modifier indicating possibility or uncertainty (cf. Li 1992: 131-134):

(74) a. *He found **something*** (specific/*non-specific) *in the box.*
b. ***Something*** (specific/*non-specific) *happened.*
c. ***Someone*** (specific/*non-specific) *is coming this way.*
d. ***Something*** (specific/*non-specific) *is floating in the water.*

(75) a. *He probably bought **something*** (non-specific) *in that store.*
b. *Apparently, **someone*** (non-specific) *is coming.*

(76) Russian

a. *Ja tebe* ***koe-što/*što-nibud*** *prinesla*.
I you-dat something bring-PAST
'I brought you **something** (specific known).' (Haspelmath 1997: 40)

b. ****Kto-nibud/kto-to*** *približaetsja*.
someone approaches
'**Someone** is approaching.' (Eremina 2012: 9)

(77) a. *Kažetsja,* ***kto-nibud'*** *približaetsja*.
seems who-INDEF approaches
'Apparently **someone** (non-specific) is approaching. (The speaker is not sure).' (Haspelmath 1997: 40)

b. *Veroyatno*, *v korobke* ***što-nibud*** *yest*.
probably in box something is
'There is probably **something** (non-specific) in the box.'

There are also quite a few grammatical structures that allow both specific and non-specific indefinite pronouns. In such cases, the choice and interpretation of the indefinite pronoun type depends on the discursive context of the sentence and the knowledge of the speaker:

(78) a. *Tomorrow, John will go* ***somewhere*** (specific). *I don't know where.*

b. *Tomorrow, John will go* ***somewhere*** (non-specific). *He hasn't decided where yet.*

Since future clauses allow both readings, in (78), John will either go to a particular place that is however not known by the speaker (specific unknown) or go to an unspecified place (non-specific). It is possible that John has not decided where to go, or the speaker does not know if there is a specific place that John wants to visit. In a similar fashion, sentences with the verb *want* also make both readings possible:

(79) a. *Mary wants to buy* ***something*** (specific) *in that store. She said it was expensive.*

b. *Mary wants to buy* ***something*** (non-specific) *for her sister . She isn't sure what to buy yet.*

Interestingly, only the non-specific interpretation is generally possible when the speaker speaks in first person. The specific reading is unlikely due to the fact that it is illogical for the speaker to inform their interlocutor that they want something specific and refuse to reveal information about the exact identity of the referent. It seems equally strange to ask for something that we are unable to identify:[16].

(80) a. ? *I want you to bring me* ***something*** (specific unknown) *from that store. I don't know what it is.*

b. ? *I want you to bring me* ***something*** (specific known) *from that store. I won't tell you what it is though.*[17]

c. *I want you to bring me* ***something*** (non-specific) *from that store. It doesn't matter what.*

Some other grammatical contexts that allow both specific and non-specific indefinites are habitual sentences and sentences with modal verbs such as *can* or *may* (cf. Giannakidou 1995, Haspelmath 1997: 41):

(81) a. *On Monday,* ***someone*** (specific) *comes and turns on the machine. I don't know who this person is.*

b. *On Monday,* ***someone*** (non-specific) *comes and turns on the machine. They send a different person each time.*

(82) a. ***Someone*** (specific) *may come. I don't know who they are.*

b. ***Someone*** (non-specific) *may come or not. I'm not sure.*

Lastly, both specific and non-specific readings are possible in sentences with universal quantifiers (e.g. English *every*). In the example bellow, the specific reading indicates that every student is reading the same particular thing (collective reading), while the non-specific interpretation is distributive, which means that every student is reading a different thing (Haspelmath 1997, 41):[18]

[16]Compare with (53) and (58).

[17]This sentence becomes acceptable if the listener has to guess what the speaker is talking about.

[18]As noted in Eremina (2012: 50-70), specific indefinite markers may appear to receive a quasi-narrow scope reading in sentences describing habitual actions or under the scope of *every* (not all speakers of Russian seem to agree). Consider the following example:

(83) a. *Every student is reading **something*** (specific). *They are reading the same thing.*

b. *Every student is reading **something*** (non-specific). *Each student may be reading a different thing.*

Compare with similar sentences in Russian. The choice of pronouns (specific/non-specific) will depend on the meaning intended by the speaker (cf. Eremina 2012: 30, 47-48):

(84) *Russian*

a. *Každyj mal'čik budet rad esli vstretit **kogo-to/kogo-nibud'** iz svoix*
every boy be.FUT happy if meet.FUT who-INDEF from his
odnoklassnic.
girl-classmates
'Every boy will be glad if [he] will-meet **some (one)** (someone specific/non-specific) of his girl-classmates.'

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)** (specific/non-specific) of my students is always called before the dean.'

In the first sentence, the use of the non-specific marker *-nibud* indicates that it does not matter to any of the boys what female classmate they meet. In contrast, the specific unknown marker *-to* tells us that there is a particular girl that every boy wants to meet. In the second scenario, the use of the specific unknown *-to* marker means every teacher knows that there is a particular

(i) *On ochen' obshitel'nyj chelovek, on (vsegda) priglashajet **kakix-to** studentov, oni vmeste*
he very sociable person, he (always) invite-PRES some(specific) student-PL, they together
*chitajut **kakije-to** knigi.*
read-PRES some(specific) book-PL
'He is a very sociable person, he (always) invites **some students**, and they read **some books** together.' (Russian, Eremina 2012: 11)

According to Eremina (2012: 11, 64-65), the seemingly narrow-scope reading stems from the fact that sentences such as the one above may refer to multiple separate situations. Each time the meeting takes place, there may be different students and different books, but they remain specific within the frame of their particular occurrence. Eremina (2012) concludes that specific indefinite markers (*-to*) may not receive a genuine narrow-scope interpretation.

student that always gets called to the dean's office. Since the student is always the same person (a particular individual), the indefinite marker used is (*-to*). The non-specific marker *-nibud* will be used if a different student is called to the dean's office each time.

2.5. Specific and non-specific indefinites: summary

The typology of indefinite pronouns put forward in Haspelmath (1997) divides indefinites traditionally known as assertive indefinite pronouns into three types on the basis of two properties, namely specificity and knowledge of the speaker.[19] Non-specific indefinites introduce a referent category (e.g. person, thing or place), which is however not linked to any particular individual entity. In contrast, specific indefinites always refer to a particular referent (a concrete entity) of a given category. Pronouns of this type can be further divided into two types with respect to the knowledge that the speaker has of the referent. Specific unknown indefinite pronouns have referents which are not identifiable by the speaker (their identity remains unknown or uncertain), while specific known indefinite pronouns are used only when the speaker knows the exact identity of the referent (and may want to withhold that information). Of course, since the source of indefinite meaning in any indefinite pronoun is its marker, the existence of non-specific, specific unknown and specific known indefinite pronouns means that there are corresponding indefinite markers for each of the three types of indefinites. In the next section, I present a cross-linguistic comparison of indefinite pronouns to establish what indefinite marker forms are used to express the non-specific, specific unknown and specific known indefinite functions in the studied languages.

[19] I do not attempt to provide a detailed semantic analysis the properties known as specificity and knowledge of the speaker. I also do not argue that these two terms directly correspond to particular syntactic features. I use the notions of specificity and knowledge of the speaker to illustrate the differences between non-specific, specific unknown and specific known indefinite forms.

Chapter 3: The data

3.1. Indefinite pronouns: the crosslinguistic analysis

The proposed analysis of specific and non-specific indefinite pronouns takes into account data from the following 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, Macedonian, Maltese, Mandarin Chinese, Mi'gmaq, Nanay, Ossetic, Persian, Polish, Portuguese, Quechua, Romanian, Russian, Serbian/Croatian, Slovene, Colombian Spanish, Swahili, Swedish, Turkish, Veps and Yakut. The data were collected and analyzed in order to establish the lexical forms of the non-specific, specific unknown and specific known indefinite markers in each language. The identified indefinite marker forms are the basis for the analysis presented in Chapter 5.

The first major source of relevant data was Haspelmath (1997), which presents a large-scale cross-linguistic analysis of indefinite pronouns. Additional data were collected and verified on the basis of a variety of linguistic sources.[20] Wherever possible, the data found in the literature were discussed with native speakers for further confirmation[21]. In a few cases, the functional distribution of indefinite forms as described in Haspelmath (1997) was not fully confirmed by other written sources or native speakers. I address these matters in Section 3.6.[22]

[20] Sources such as articles, dictionaries and descriptions of grammar were used.

[21] During interviews or in questionnaires, native speakers were presented with sets of sentences and asked to translate them. For each of the three analyzed indefinite pronoun types, I chose sentences that allowed only one particular reading of the indefinite pronoun. Contexts were provided to help interviewed speakers to imagine specific situations. Whenever necessary, native speakers were asked about their judgments and interpretation concerning the discussed sentences. Further questions were asked if there were multiple possible translations.

[22] A full list of examples and their sources is provided in the appendix.

3.2. Three separate markers: Russian and Lithuanian

As shown in Chapter 2, Russian is a language in which the non-specific, specific unknown and specific known indefinite functions are all expressed by completely different markers. The three markers are *-nibud* (non-specific), *-to* (specific unknown) and *koe-* (specific known). Lithuanian is another language in which the three indefinite functions have separate corresponding indefinite markers. These markers are -nors (non-specific), *kaž-* (specific unknown) and kai- (specific known). The sentences in (85) and (86) show the use of the non-specific, specific unknown and specific known indefinites in Lithuanian and Russian:

(85) Lithuanian (Haspelmath 1997: 276, Kozhanov 2015: 467, Pilka 1984: 29 as cited in Haspelmath 1997: 40)

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

b. *Gali* ***kaž-kas*** *ateiti*.
can INDEF-who come
'**Someone** (specific unknown) can come.'

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.'

(86) Russian (Eremina 2012: 7, 21, 74)

a. *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.'

b. *Masha prigotovit* ***što-to*** *vkusnoje na uzhin*.
Masha cook-FUT what-INDEF delicious for dinner
'Masha will cook something (specific unknown) delicious for dinner.'

c. *Ja tebe* ***koe-što*** *prines*. *Smotri*, *kakaja dynia*.
I you-DAT INDEF-what brought. Look which melon.
'I brought you **something** (specific known). Look at this melon.'

Table 1 shows the non-specific, specific unknown and specific known indefinite markers in Russian and Lithuania:

Table 1. Indefinite markers in Lithuanian and Russian.

	non-specific	specific unknown	specific known
Lithuanian	-nors	kaž-	kai-
Russian	-nibud	-to	koe-

3.3. Single indefinite marker

The majority of languages in the studied languages sample express the non-specific, specific unknown and specific known indefinite functions with the same phonological form. This means that non-specific, specific unknown and specific known indefinite pronouns in those languages cannot be distinguished on the basis of their form, and their interpretation depends on the context. These languages are: Bulgarian, Catalan, Czech, Dutch, English, French, Slovak, German, Hausa, Hebrew, Hindi, Hungarian, Icelandic, Italian, Japanese, Kazakh, Classical Greek, Korean, Latvian, Lezgian, Maltese, Polish, Portuguese, Romanian, Serbian/Croatian, Colombian Spanish, Swedish and Turkish. The following examples show data from English, Polish, Japanese, Korean and Lezgian (Northeast Caucasian):

(87) English

a. *Let's go **somewhere*** (non-specific).

b. *There is **someone** behind the door* (specific unknown).

c. *I have **something** for you* (specific known).

(88) Polish

a. *Przynieś mi **co-ś** do jedzenia.*
bring.impv I.DAT what.ACC-INDEF to eating.GEN
'Bring me **something** (non-specific) to eat.'

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** called. Guess who.'

(89) Japanese (Haspelmath 1997: 312)

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

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

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

(90) Korean (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.'

(91) Lezgian (Haspelmath 1997: 296)

a. *Za wa-z* ***sa wuč-jat'ani*** *luhu-da.*
I.ERG you-DAT one what-INDEF say-FUT
'I'll tell you **something** (specific known).'

b. *Gilač ar-ar* ***wučiz-jat'ani*** *xükwe-zma-č*.
now letter-PL why-INDEF come-CONT.IMPF-NEG
'Now the letters don't come anymore **for some reason** (specific unknown).'

c. *Sa* ***ni-z-jat'ani*** *ewer ce*!
one who-DAT-INDEF call give.IMPV
'Call **someone** (non-specific)!'

Table 2 shows markers corresponding to the non-specific, specific unknown and specific known indefinite functions in English, Polish, Japanese, Korean and Lezgian. In these languages, markers representing the three indefinite functions have the same phonological form:

Table 2. Indefinite markers in English, Polish, Japanese, Korean and Lezgian.

	non-specific	specific unknown	specific known
English	some-	some-	some-
Polish	-ś	-ś	-ś
Japanese	-ka	-ka	-ka
Korean	-nka	-nka	-nka
Lezgian	-jat'ani	-jat'ani	-jat'ani

3.4. Ossetic, Georgian and Yakut

In Ossetic, Georgian and Yakut, the specific known and unknown functions are expressed by a single indefinite marker to the exclusion of the non-specific function, which has a separate indefinite marker. In Georgian the markers are *-γac*, for the specific functions (known and unknown), and *-me*, for the non-specific function (Hewitt 1996: 68). In Ossetic, the markers are *-dær* (specific known/unknown) and *-ty* (non-specific). Lastly, in Yakut, the markers are *-ere* (specific known/unknown) and *-eme* (non-specific). The following sentences show the use of specific and non-specific indefinites in Ossetic, Georgian and Yakut:

(92) Ossetic (Haspelmath 1997: 281, Kulaev 1958: 52 as cited in Haspelmath 1997: 281)

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.'

(93) Georgian (Haspelmath 1997: 304, Vogt 1971: 47)

a. *Es c'igni* ***sad-γac*** *v-išove.*
this book where-INDEF 1SG-found
'I found this book **somewhere** (I could say where).'

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

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

(94) Yakut (Haspelmath 1997: 290, Afanas'ev and Xaritonov 1968: 250 and Ubrjatova 1982: 200 as cited in 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).'

The non-specific, specific unknown and specific known functions in Georgian, Ossetic and Yakut are expressed by the following markers. The specific known and specific unknown markers have the same phonological form:

Table 3. Indefinite markers in Georgian, Ossetic and Yakut.

	non-specific	specific unknown	specific known
Georgian	-me	-γac	-γac
Ossetic	-ty	-dær	-dær
Yakut	-eme	-ere	-ere

Nanay (Tungusic) also seems to belong to this category of languages. The non-specific marker is *-daa*, while *-nuu* appears to be the specific indefinite marker (known and unknown). However, as mentioned in the source of the data (Haspelmath 1997: 66-67), the data presented therein are incomplete. In consequence, I do not have detailed information concerning the specific unknown and specific known functions, which means that the actual functional distribution of *-nuu* may be different:[23]

(95) Nanay (Onenko 1986: 441-442, 449 as cited in Haspelmath 1997: 66-67)

a. ***Uj-nuu*** *žook-či* *laŋ* *ži-či-ni*.
who-INDEF house-DIR near go-PAST-3SG
'**Someone** (specific unknown/known) 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 unknown/known).'

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.'

3.5. Latin

In Latin the non-specific and specific unknown functions are represented by the same form *ali-* (Haspelmath 1997: 254):

[23] Due to the incompleteness of the data, I will omit Nanay in the summary.

(96) Latin (Haspelmath 1997: 254)

a. *Tetigit me **ali-quis**.*
touched me INDEF-who
'**Somebody** (specific unknown) hath touched me.' (New Testament, Luke 8:46)

b. *At ille intendebat in eos, sperans se **ali-quid** accepturum ab eis.*
but that gave.heed in them hoping self INDEF-what accept.FUT from them
'And he gave heed unto them, expecting to receive **something** (non-specific) of them.' (New Testament, Acts 3:5)

To speak about a specific referent which is known to the speaker but its exact identity is not mentioned, the marker *-dam* is used (Gianollo 2018: 49-58, Haspelmath 1997: 254):[24]

(97) New Testament, Mark 9:38

*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.'

The indefinite marker forms in Latin are as follows. The non-specific and specific unknown markers have the same phonological form to the exclusion of the specific known marker:

Table 4. Indefinite markers in Latin.

	non-specific	specific unknown	specific known
Latin	ali-	ali-	-dam

3.6. Incomplete paradigms

There are a number of languages in the studied language sample in which indefinite marker paradigms are incomplete. In other words, in a number of languages indefinite markers (and corresponding indefinite pronoun forms) for one or more of the three indefinite functions are not available. A further discussion of such languages and the phenomenon of paradigm gaps

[24] See also Orlandini (1981).

can be found in Section 6.2. Table 5 presents the languages from the studied sample without one or more of the three studied indefinite marker types:[25]

Table 5. Incomplete paradigms.

	non-specific	specific unknown	specific known
Kannada	-aadaruu	-oo	–
Veps	- ni	-ni	–
Quechua	-pis/-pas	-chi/-cha	–
Mandarin Chinese	wh-pronoun	–	–
Irish	–	–	–
Swahili	–	–	–
Filipino	–	–	–

3.6.1. Kannada

Kannada has only two indefinite pronoun series, namely the non-specific *-aadaruu* series and the specific unknown *-oo* series. According to the available data, that is Bhat (1981) and Haspelmath (1997: 305-306), specific known indefinite pronouns appear to be absent from the language. In cases where the identity of the referent is known, the speaker may use a generic noun with the modifier *ondu* 'one':

(98) Kannada (Haspelmath 1997: 306-307)

a. *Raamuvige **ondu** pustaka beekaagide.*
Ramu-to one book want-having-is
'Ramu wants a (specific known) book.'

b. ***Yaar-oo** bandaru.*
someone came
'*Someone* (specific unknown) came.'

c. *Raamuvige **een-oo** sigalilla.*
Ramu-to what-INDEF got-not
'Ramu did not get **something** (specific unknown).'

[25]The sources of the data are provided in the following sections.

d. ***Ellig-aadaruu*** *hoogu*.
where-INDEF go
'Go **somewhere** (non-specific).'

The specific unknown marker *-oo* may not be used to express the specific known function, as seen in the example below:

(99) Kannada (Bhat 1981: 7)

Yaar-oo *bandaru* (**yaaru uuhisi*).
who-INDEF came who guess
'**Someone** came (*guess who).'

3.6.2. Veps

In Veps, a language closely related to Finnish, only the specific unknown and non-specific functions may be expressed with regular indefinite pronoun series. According to Hienonen (2010), specific known pronouns are absent from the language with the sole exception of the pronoun *eraz* 'someone/some/other', which can appear in the non-specific, specific unknown and specific known contexts. *Eraz* is not a part of any indefinite pronoun series and is the only pronoun of its kind. The pronoun appears to share its origins with the Finnish *eräs* 'a certain/some/one', which suggests that perhaps it should be considered similar the English generic pronoun *one*. For these reasons, I will not consider *eraz* to be a pronoun of the assertive indefinite category:

(100) Veps (Setälä and Kala 1951: 457 as cited in Hienonen 2010: 286)

E-n ota, *minä* ***erase-n*** *spolĺubi-n*, ***erase-n*** *mehe-le*
NEG-1SG take.CON I some-GEN have (a passion)-1SG some-GEN husband-ALL
ota-n.
take-1SG
'I don't take (her), I have a passion for **someone** (specific known), I will marry **someone** (specific known).'

The basic indefinite pronoun series is the *-ni* series which is used in the specific unknown and

non-specific functions:[26]

(101) Veps (Kodima 1999/11-12: 2, Markan evangelii 1992: 8: 34 as cited in Hienoen 2010: 287)

a. *Kuziže-n taga-päi tule-b **mitte-ni** sarnaline olii.*
spruce-GEN behind-ELA come-3SG some-INDEF fairy-tale creature
'**Some** (specific unknown) fairy-tale creature comes out from behind the spruce.'

b. *Ku **ken-ni** tahtoi-b elä-da minu-n mödhe, se unohta-g-ha*
if who-INDEF want-3SG live-1INF I-GEN along it forget-IMP-3SG
iče-ze-n pol-he.
self-POSS.3SG-GEN side-ILL
'If **anyone** wants to live like me, he must forget himself.'

3.6.3. Quechua

Quechua is also a language which lacks one of the three types of assertive indefinite pronouns. The collected data indicates that only two indefinite pronoun series are available in Quechua, namely the non-specific *-pis*/*-pas* series and the specific unknown *-chi*/*-cha* series:[27]

(102) Quechua specific unknown indefinites (Cusihuáman 2001: 234, Haspelmath 1997: 310)

a. ***Ima-ta-chi** wambra yurapa waqta-n-chaw riqa-rqa-n.*
what-ACC-VAL boy tree behind-3SG-LOC see-PAST-3SG
'The boy saw **something** (specific unknown) behind the tree.'

b. ***Pi-wan-chi** qanyan awtobus-chaw parla-rqu-u.*
who-COMIT-VAL yesterday bus-LOC talk-PAST-1SG
'I talked to **someone** (specific unknown) on the bus yesterday.'

c. ***Pi-cha** haqay-ta hamu-sha-n!*
who-CONJ over.there-ACC come-PROG-3
'**Someone** unknown is coming over there./I don't know who is coming over there.'

[26]The only example given in Hienonen (2010) for the specific unknown function contains the form *mitte-ni* (some/something) used as an indefinite modifier.

[27]The forms vary depending on the dialect.

(103) Quechua non-specific indefinites (Cusihuáman 2001: 107, Haspelmath 1997: 310)

a. ***Ima-ta-pis*** *rika-rnin-qa willa-ma-y.*
what-ACC-INDEF see-2SG-TOP tell-1SG-IMPV
'If you see **anything** (something non-specific), tell me.'

b. *Mana musya-pti-iki-qa* ***pi-ta-pis*** *tapuku-y.*
not know-CONV-2SG-TOP who-ACC-INDEF ask-IMPV
'If you don't know, ask **somebody** (non-specific).'

c. *Dueño-n ni-n nisita-n=si* ***pi-lla-ta=pas*** *kay-pi tiya-y-ta.*
owner-3 say-3 need-3=REP who-LIM-ACC=ADD this-LOC sit-INF-ACC
'The owner says that he needs **someone** (non-specific) to live here.'

The *-pis*/*-pas* series of indefinites is consistently mentioned in literature on the Quechua language such as Anchorena (1874), Cusihuáman (2001), Faller (2020), Haspelmath (1997: 310-311), Parker (1976), Shimelman (2017), Weber (1989). The second kind of indefinite expressions, i.e. the *-chi*/*-cha* series, is not so well attested. I have found a reference to it only in Faller (2019, 2020) and Cusihuáman (2001: 234-235) (Cuzco Quechua). Both sources mention the use the conjectural suffix *-cha* to create indefinite forms which refer to things, people, places, etc. that the speaker does not know. Weber (1989) (Huallaga Quechua) also mentions the suffix *-chi*, which indicates that a statement is a conjecture on the part of the speaker. This source however makes no mention of indefinite pronoun forms. Lastly, an interesting case is the account found in Haspelmath (1997: 310-311) since it describes specific unknown indefinites in Quechua as bare wh-pronouns. However, I was not able to find a similar description of Quechua indefintes in any other source. Additionally, the examples provided in Haspelmath (1997: 310-311) (Ancash Quechua) clearly show wh-forms marked with the suffix *-chi*, which is glossed as VAL (validational). This validational *-chi* suffix appears to be the same as the conjectural *-chi* suffix found in Weber (1989) and the conjectural *-cha* suffix mentioned in Faller (2019) and Cusihuáman (2001: 234-235) (cf. Frawley 2003: 431-432). I will follow Faller (2019, 2020) and Cusihuáman (2001) and consider the *-chi*/*-cha* suffix to be the specific unknown indefinite marker in Quechua.

As for the specific known function, none of the sources mention any indefinite pronoun forms used in this function. It is possible that if the identity of the referent is known to the speaker, they may use a bare noun with the modifier *huk* 'one' (Espinoza 1997: 16, Faller 2020:

23):

(104) Espinoza (1997: 16)

”*May-manta=n ka-nki-chis?*” *ni-spa tapu-wa-n* ***huk runa***.
where-ABL=BPG be-2-PL say-NMLZ.SS ask-10-3 one man
‘“Where are you from?” **a man** asked me.’

The data currently available to me are however not sufficient to clearly and conclusively establish what forms are used to refer to entities that are considered specific and known by the speaker.

3.6.4. Mandarin Chinese

In Mandarin Chinese, only one of the three assertive indefinite pronoun types is available, namely non-specific indefinites which take the form of bare wh-pronouns (e.g. *shénme* ‘what/something’ and *shéi* ‘who/somebody’):

(105) Mandarin Chinese (Li 1992: 152, Lin 1998: 3)

a. *Chī, diǎn* ***shénme*** *zài zǒu ba*!
eat a.bit what then go PT
‘Please eat a little **something** (non-specific) before you leave.’

b. *Keneng/xiangbi* ***shei*** *you qifu ta le*
possibly/presumably who again bully him ASP
‘Possibly/presumably, **somebody** bullied him again.’

c. *Wǒ xiǎng hē diǎnr* ***shénme***.
I want drink a.bit what
‘I want to drink **something** (non-specific).’

There are no specific indefinite pronouns in Chinese. In contexts where the referent is interpreted as specific unknown or specific known, speakers will use generic nouns such as *rén* ‘person’ or *dōngxi* ‘thing/stuff’. Such nouns often appear in combination with the existential verb *yǒu* ‘there is’:

(106) Mandarin Chinese (Haspelmath 1997: 308)

a. *Yǒu* ***rén*** *dǎ diànhuà le. Wo bù zhīdao shì shéi.*
exist man hit phone PERF I not know is who
'**A man** called. I don't know who (specific unknown context).'

b. *Yǒu* ***rén*** *dǎ diànhuà le. Cāi yī cāi shì shéi.*
exist man hit phone PERF guess one guess is who
'**A man** called. Guess who (specific known context).'

c. *Wǒ yào gěi nǐ yīgè* ***dōngxī****. Cāi cāi kàn shì shénme.*
I want give you one-CL thing guess guess see is what
'I want to give you **a thing**. Guess what it is.'

Chinese generic nouns used in contexts in which we would expect to see specific indefinite pronouns do not seem to exhibit any special properties that would indicate that they should be interpreted as indefinite pronouns, for example, they may appear together with demonstratives, numerals and classifiers. It is likely that generic nouns and existential constructions with generic nouns simply serve as strategies compensating for the lack of specific indefinite pronouns.

3.6.5. Finnish, Greek and Basque

According to the data found in Haspelmath (1997: 293-294), languages such as Finnish, Greek and Basque should also belong to the category of languages without all three assertive indefinite pronoun types. All three languages are described as lacking indefinites used in the specific known function. I was however unable to confirm the data presented in Haspelmath (1997) with other sources and native speakers' judgments. Below, I discuss the collected data. First, consider the data from Finnish presented in Haspelmath (1997); the same indefinite forms are used in the specific unknown and non-specific functions:

(107) Finnish Haspelmath (1997: 293-294)

a. *****Joku*** *soitt-i. Arvaa kuka (se oli)?*
someone call-PAST(3SG) guess:IMPV who it was
'**Someone** (specific known) called. Guess who it was.'

b. ***Joku*** *soitt-i*, *mutta en* *saa-nut* *nime-stä*
someone call-PAST(3SG) but NEG:1SG get-PAST.PTCP name-ELA
selvää.
clear:PRTV
'**Someone** (specific unknown) called, but I didn't understand the name.'

c. *Hän haluaa mennä naimisiin* ***jonkun*** *kanssa*, *jolla* *on tumma tukka*.
she wants go marriage someone with who-on is dark hair
'She wants to marry **someone** (non-specific) with black hair.'

As mentioned in Haspelmath (1997: 293-294), the *joku*/*jokin* 'someone/something' series should be considered correct only in the non-specific and specific unknown functions. To refer to an entity that is specific known, a speaker may use the modifier *eräs* 'certain' with a generic noun (Haspelmath 1997: 293, Korpela 2015). A similar account is given in White (2006), which mentions that the *joku*/*jokin* forms refer to referents unknown to both the speaker and the listener. Modifiers such as *eräs* 'certain' and *yksi* 'one' can be used together with generic nouns when the listener knows the identity of the referent.

The unavailability of the *joku*/*jokin* series in the specific known function was however not confirmed by native speakers. According to two different native sources, pronominal forms are also correct in the specific known function, which means that example (107-a) should not be considered ungrammatical. The following example from a native speaker shows how *jokin* 'something' is used in the specific known function:

(108) *Arvaa mitä*, *minulla* *on sinulle* ***jotain*** *todella suurta kerrottavaa*.
guess what have.1p.sg be for.you something really great tell-INF
'Guess what, I have **something** (specific known) really great to tell you.'

Therefore, I will include Finnish in the category of languages in which the non-specific, specific unknown and specific known indefinite functions are expressed by the same set of pronouns.

Greek is supposed to be another language without specific known indefinite pronouns. According to Haspelmath (1997: 265-267), the *ká-* series should appear only in the non-specific and specific unknown functions. The use of the *ká-* series in the specific known functions is described as ungrammatical:

(109) Greek (Haspelmath 1997: 266-267)

a. ***Ká-pjos*** *tilefónise. Mándepse pjos*!
INDEF-who phoned guess:IMPV who
'**Someone** (specific known) called. Guess who!'

b. ***Ká-pjos*** *tilefónise. Dhen kséro pjos.*
INDEF-who phoned not know.1p.sg who
'**Someone** called (specific unknown). I don't know who.'

c. *Fére* ***ká-ti*** *na fáme*!
bring:IMPV INDEF-what SBJV eat.2p.pl
'Bring **something** (non-specific) to eat!'

As in the case of Finnish, the data concerning the specific known function presented in Haspelmath (1997) were not confirmed by native speakers. In other words, the use of the *ká-* series in the specific known function was considered grammatical.

Haspelmath (1997: 315-316) also mentions Basque as a language without specific known indefinite pronouns. According to the data provided in Haspelmath (1997: 315), the word *bat* 'one' is used to refer to specific known entities:

(110) Haspelmath (1997: 316)

Kanpoa-n zengoze-n-en-ean ***baten*** ***batek*** *dei egin zizun. Esan nor.*
away-LOC be-2SG-REL-LOC one:GEN one:ERG call make he:you:has say who
'**Someone** called while you were away. Guess who it was.'

Indefinite pronouns with the suffix *-bait* are used in the non-specific and specific unknown functions:

(111) Basque (Saltarelli 1988 as cited in Haspelmath 1997: 316)

a. ***Nor-bait-i*** *utzi nion argazki-makina eta ez naiz oroi-tzen*
who-INDEF-DAT lend I.it.to.him photo-machine and NEG I.it remember-HAB
nor-i.
who-DAT
'I lent the camera to **someone** (specific unknown) and I do not remember to

whom.'

b. ***Nora-bait*** *joa-te-ko gogoa dut, baina ez dakit nora.*
where-INDEF go-HAB-REL thought I.have.it but NEG I.know.it where
'I have an urge to go **somewhere** (non-specific) but I do not know where.'

The description of Basque indefinite pronouns found in Haspelmath (1997: 315) is not confirmed by other sources on the language. For example, grammars of Basque such as Patrick and Ibarrondo (2001) and de Rijk (2008) do not mention that the *-bait* series should be considered ungrammatical in the specific known function. Additionally, de Rijk (2008: 692) provides an example which indicates that *-bait* indefinites may be available in the specific known function:

(112) de Rijk (2008: 692)[28]

Hiretzat ere aukeratuta neukan ***norbait***.
for.you too chosen had.1SG someone
'For you too I had chosen **someone**.'

The discussed conflicting descriptions in the collected data may stem from inadvertent inaccuracies in either Haspelmath (1997) or other sources. At the same time, it is also possible that the differences are caused by variation within the discussed languages. Perhaps, some native speakers of Finnish, Greek and Basque have access to fully functional specific known indefinite pronouns, while other will consider the use of indefinite pronoun in the specific known function ungrammatical. However, since at this point I am unable to present a conclusive answer to this issue, I will follow the speaker' judgments and place Finnish and Greek in the same category as English. As for Basque, my data is currently inconclusive.

3.6.6. Irish, Swahili and Filipino

Irish, Swahili and Filipino are languages without non-specific, specific unknown and specific known indefinite pronouns.[29] Due to the lack of such pronouns, these languages employ a

[28] Glosses mine, JD.

[29] Assertive indefinite pronouns are also reported to be absent from Lango (Noonan 1992: 166). Generic nouns are used instead:

(i) Lango

number of replacement strategies that give rise to interpretations that may be considered close to the functional semantics of indefinties. The forms used in place of indefinite pronouns include generic nouns, generic nouns with modifiers and existential constructions.[30]

In Irish, indefinite entities are referred to with the use of generic nouns or nouns with the indefinite modifier *éigin* 'some/certain'. To my knowledge, Irish generic nouns that appear in contexts where indefinite pronouns would be expected are not grammaticalized and cannot be interpreted as indefinites (cf. Windisch 1882: 64):

(113) Irish (Haspelmath 1997: 279, Bhaldraithe 1959: 302-303)

a. *Tá* ***rud*** *agam le rá leat.*
   is thing on:me for telling to:you
   'I have **something** (lit. thing) to tell you.'

b. *Dúirt* ***duine éigin*** *liom é.*
   told person certain to:me he
   '**Somebody** (lit. a certain person) told me.'

c. *Abair* ***rud éigin****.*
   say:IMPV thing certain
   'Say **something** (lit. some thing).'

Swahili is another language without specific and non-specific indefinite pronouns. As in Irish, speakers use regular generic nouns in contexts where indefinite pronouns would be used in other languages. Such nouns are not grammaticalized forms and should not be considered indefinites. Additionally, generic nouns in Swahili may also be accompanied by the modifier *fulani* 'certain' (cf. Mpiranya 2015):

Ànénò ŋat.
1SG-see-PERF person
'I saw someone (lit. 'a person').'

I will not discuss Lango in a separate section since my data consists of a single example from Noonan (1992: 166).

[30] In the examples below I do not use words such as *someone* or *something* as direct translations.

(114) Swahili (Haspelmath 1997: 302-303)

a. ***Mtu*** *a-li-ni-gusa*.
man 3SG-PAST-me-touch
'**Somebody** has touched me.'

b. *Yesu a-li-mw-ambia kwamba a-wa-pe maskini* ***kitu***.
Jesus 3SG-PAST-him-tell that 3SG-them-give:SBJV poor thing
'Jesus told him to give **something** to the poor.'

Filipino is a language in which we also observe the absence of specific and non-specific indefinite pronouns. In order to refer to entities deemed indefinite, speakers may use generic nouns or generic nouns with the modifier *isang* 'one' (which also serves as the indefinite article). It is also very common to use existential constructions with *may*/*mayroon* 'there is':

(115) Filipino

a. *Dalhan mo ako ng makakain*!
bring.IMPV you me of edible
'Bring me **something** to eat!'

b. *Meron tumawag pero di ko alam kung sino*.
there.is called but NEG I know if who
'**Someone** called, but I don't know who.'

c. *Meron akong ibibigay para sa iyo*. *Hulaan mo kung ano*.
there.is I give.FUT for yours guess.IMPV you if what
'I have **something** to give you. Guess what.'

3.7. Indefinite pronouns: syncretism

The cross-linguistic analysis of indefinite pronouns reveals that the non-specific, specific unknown and specific known indefinite functions may be expressed with one, two or three indefinite pronoun series. This also means that these indefinite functions may be represented by one, two or three indefinite markers. Subsequently, on the basis of the collected data, it is possible to group languages into four categories:

(116) a. Languages with only one indefinite marker for the three indefinite functions (e.g. English, Polish, Japanese, etc.)

b. Languages with three separate indefinite markers for the three indefinite functions (e.g. Russian and Lithuanian)

c. Languages where the specific unknown and specific known functions are expressed with the same marker to the exclusion of the non-specific function (e.g. Ossetic, Yakut, Georgian)

d. Languages where the specific unknown and non-specific functions are expressed with the same marker to the exclusion of the specific known function (e.g. Latin)

As argued in Haspelmath (1997), the fact that in a language multiple indefinite functions may share a single phonological form (indefinite marker) is not a matter of coincidence but stems from a regularity that can be understood and described. Having conducted an analysis of indefinite pronouns across multiple languages, Haspelmath (1997) proposes the following map of indefinite functions:[31]

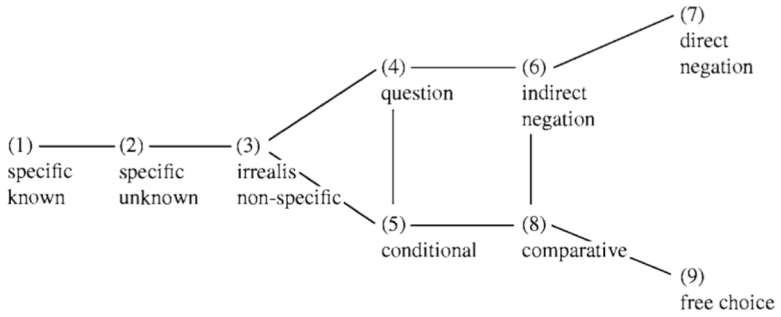


Figure 2. The implicational map for indefinite pronoun functions. (Haspelmath 1997: 4).

On the basis of the conducted analysis, Haspelmath (1997) makes a particular prediction concerning indefinite functions and their overt realizations (indefinite pronouns and markers). It is proposed that an indefinite pronoun series, and consequently its corresponding indefinite marker, may express multiple indefinite functions, but only as long as those functions constitute adjacent elements on the map shown in Figure 1.

The regularities described in Haspelmath (1997) are not exclusive to the category of indefinite pronouns and should be viewed in the context of a certain cross-linguistically common phenomenon. This phenomenon is known as syncretism and can be understood as phonological

[31] See also Haspelmath 2003.

conflation of two or more items in a paradigm (cf. Caha 2009: 6). Examples of syncretism may be found across a variety of grammatical domains, for instance in case paradigms. The following table shows syncretic forms in the case paradigms of selected nouns in Polish. As shown, noun forms corresponding to different case inflections may be phonologically identical:

Table 6. Case syncretism in Polish.

case	dog.SG	window.SG	hundred.SG
NOM	pies	okn-o	sto
ACC	ps-a	okn-o	sto
GEN	ps-a	okn-a	stu
DAT	ps-u	okn-u	stu
INS	ps-em	okn-em	stu
LOC	ps-ie	okn-ie	stu

As argued in Bobaljik (2007, 2012), syncretism may target only contiguous cells in a paradigm. In other words, in an ordered set of forms we should never see items that are syncretic but do not occupy adjacent positions in the set. This regularity is known as the *ABA generalization.

Taking into account the data presented in this chapter and the *ABA generalization, I will argue that the implicational map of indefinite pronoun functions proposed in Haspelmath (1997) shows the (relative) order of items in the paradigm of the non-specific, specific unknown and specific known indefinite markers.[32] When arranged in accordance with the order shown in Figure 2, assertive indefinite marker forms found cross-linguistically will match one of three patterns of syncretism (AAA, AAB, ABB). At the same time, the ABA pattern, that is when the non-specific marker shares its phonological form with the specific known marker to the exclusion of the specific unknown marker, will remain unattested. The following table shows the attested forms of indefinite marker syncretism:

[32]Note that the data collected for the purpose of this analysis are limited to indefinite pronouns corresponding to the non-specific, specific unknown and specific known indefinite functions. I will not discuss other functions shown on the implicational map.

Table 7. Indefinite pronoun paradigm: patterns of syncretism.

pattern	non-specific	specific unknown	specific known
AAA	■	■	■
ABB		■	■
AAB	■	■	
*ABA	■		■

The cross-linguistic analysis of indefinite pronouns presented in Haspelmath (1997), as well as the comparative study conducted for the purpose of this analysis provide no examples that contradict the order of the non-specific, specific unknown and specific known functions shown on the implicational map and the *ABA generalization. Therefore, the analyzed data sample can be considered to constitute considerable evidence supporting the idea that the non-specific, specific unknown and specific known indefinite functions and their corresponding indefinite markers form a particular ordered set in which syncretism may arise, or in other words, a paradigm. The following table presents the summary of the conducted data analysis:

Table 8. Indefinite pronoun paradigms: data.

	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	-jat'ani	-jat'ani	-jat'ani	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	xi-	xi-	xi-	AAA
Hungarian	vala-	vala-	vala-	AAA

Hebrew	-šehu	-šehu	-šehu	AAA
Turkish	bir-	bir-	bir-	AAA
Latvian	kaut-	kaut-	kaut-	AAA
Catalan	algú	algú	algú	AAA
Dutch	iets	iets	iets	AAA
Finnish	joku	joku	joku	AAA
French	quelqu'un	quelqu'un	quelqu'un	AAA
German	jemand	jemand	jemand	AAA
Greek	ká-	ká-	ká-	AAA
Mi'gmaq	nat-	nat-	nat-	AAA
Hausa	wani	wani	wani	AAA
Hindi	koii	koii	koii	AAA
Icelandic	ein-	ein-	ein-	AAA
Italian	qualcuno	qualcuno	qualcuno	AAA
Kazakh	bir-	bir-	bir-	AAA
Slovene	ne-	ne-	ne-	AAA
Classical Greek	tis	tis	tis	AAA
Portuguese	alguém	alguém	alguém	AAA
Colombian Spanish	alguien	alguien	alguien	AAA
Swedish	någon-	någon-	någon-	AAA
Yakut	-eme	-ere	-ere	ABB
Georgian	-me	-γac	-γac	ABB
Ossetic	is-	-dær	-dær	ABB
Latin	ali-	ali-	-dam	AAB
Russian	-nibud	-to	koe-	N/A
Lithuanian	-nors	kaž-	kai-	N/A
unattested				ABA

In the next section, I will discuss the core assumptions and methodological tools of a generative framework known as nanosyntax. Over the years, the nanosyntactic model of grammar and its system of phrasal spellout have been used multiple times to analyze the fine-

grained structure of different domains of grammar and explain the emergence of syncretism in those domains. Nanosyntactic works dealing with syncretism and the *ABA generalization include studies and analyses of grammatical domains such as case (Caha 2009, Türk and Caha 2021), spatial adpositions (Pantcheva 2011, Tolskaya 2018), participles (Starke 2006, Taraldsen Medová and Wiland 2018, Caha and Taraldsen Medová 2020), verbs (Jabłońska 2007, Taraldsen Medová and Wiland 2019), negation (De Clercq 2013, 2020), numerals (Wągiel and Caha 2021), demonstratives (Lander and Haegeman 2016), noun class markers (Taraldsen 2010, Taraldsen et al. 2018), complementizers (Baunaz and Lander 2018b) and wh-pronouns (Wiland 2018, 2019). I will show that assertive indefinite pronouns constitute another domain of grammar that can be analyzed under the methodological principles of nanosyntax.

Chapter 4: Nanosyntax

4.1. Nanosyntax: core tenets

Nanosyntax (Starke 2009, among others) is a theory of grammar which aims to reveal the complexity and intricacies of syntactic structures through precise and meticulous analysis of linguistic data. This endeavor stems from the simple fact that over the years the continuous study of language has revealed that grammar is much more complex and robust than it was previously assumed. In consequence of such findings, we have to propose more elaborate models of the internal structure of language if we are to accurately and comprehensively represent all its details. Some of the most significant steps towards understanding the fine-grained nature of syntactic structure include Belletti (2004), Cinque (1999, 2002), Cinque and Rizzi (2008) and Rizzi (1997, 2004).[33] It was these works (among other) that gave rise to what is known as the cartographic approach. Not only does the nanosyntactic framework share its main goal with cartography, that is to create comprehensive and precise representations of syntactic structures (cf. Rizzi 2013: 1), but also draws some of its core tenets from cartographic studies.

The first core tenet which nanosyntax inherited from the cartographic endeavor is the one feature - one head maxim (OFOH) (Cinque and Rizzi 2008: 50, Kayne 2005). In consequence of OFOH, every syntactic terminal node is considered to constitute only a single syntactic feature (Starke 2011, 2014, see also Caha 2020, 2021).[34] In nanosyntax, this assumption essentially blurs the boundaries between syntax and morphology due to the fact that single-feature heads are smaller than words or even morphemes.[35] In other words, the form of lexical items (words and morphemes) does not matter since they are just surface realizations of structure derived in syntax. For example, a pronoun such as English *he* contains number, gender, person and case

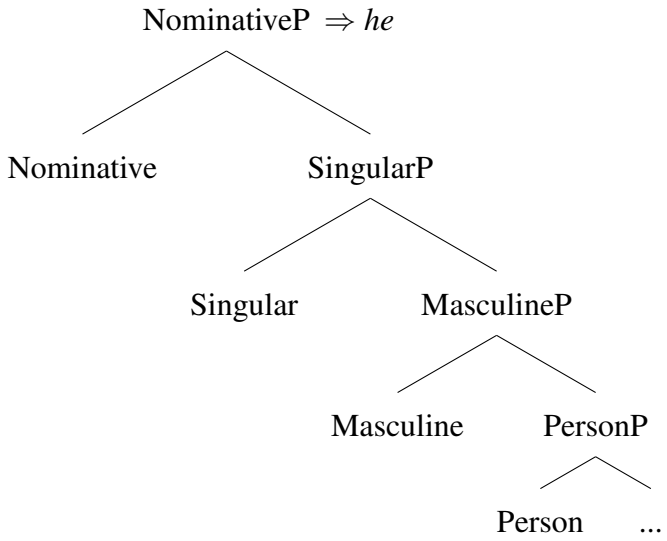
[33] Of particular importance for the development of the cartographic approach were the analyses of the left periphery presented in Rizzi (1997, 2001, 2004), Aboh (2004), Belletti (2004) and Haegeman (2006, 2012) and the functional positioning of adverbs proposed in Cinque (1999).

[34] Features are also considered privative, which means that they are not assigned +/- values.

[35] Since in the nanosyntactic model, morphology and syntax are considered a single component of language, the term morphosyntax is used.

features. All these features are not an unorganized bundle but correspond to separate projecting heads and have to be arranged into a structure (hierarchy) by the syntactic component:

(117) English pronoun *he*



As shown in (117), the lexical item (phonological exponent) *he* is the surface realization of a piece of syntactic structure. The features that are lexicalized as *he* had to be assembled as separate heads into an ordered structure (i.e. syntactic constituent) inside the syntactic module (Starke 2011, 2014, Caha 2020).

Another important assumption that is shared by cartography and nanosyntax is that despite the surface differences between languages, they are all in fact uniform in structure. This claim goes back to the Uniformity Principle of Chomsky (2001, 2), which states that (in the absence of counterevidence) the internal structure of languages should be considered universal.[36] Following this assumption, cartographic and nanosyntactic analyses present syntax as a computational system with a crosslinguistically universal sequence of functional projections used as the blueprint (Baunaz et al. 2018: 16-20).[37] In nanosyntax, this functional sequence (or fseq) provides provides features that are assembled as terminal nodes into syntactic structures.

[36]The exact formulation provided in Chomsky (2001) is as follows: "In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances".

[37]Evidence for the existence of a universal functional sequence comes from various domains of grammar such as the CP domain (Belletti 2004, Rizzi 1997), adverbs (Cinque 1999), adjectives (Cinque 2010), negation (Haegeman and Zanuttini 1991, Haegeman 1995), the VP structure (Hale Keyser 1993, Ramchand 2008), the nominal domain (Abney 1987, Cinque 2005, Giusti 1997), the TP domain and inflection (Cinque 1999, Pollock 1989) and more.

There are however two competing claims concerning how the universality of the functional sequence should be understood (cf. Baunaz and Lander 2018c). According to the strict approach to the fseq, an invariable number of features is projected in a fixed order in all languages (Cinque 1999, 2013). In other words, the same fseq is considered to be always projected cross-linguistically even if the presence of all postulated features is not visible in every language. This also means that if there is evidence for the existence of a particular feature (head) in one language, all languages should be assumed to have this feature as well (cf. Cinque 2006, Kayne 2005: 12, Cinque and Rizzi 2008). In consequence, languages may only differ with respect to how each language realizes the functional sequence.

The other approach to the fseq is motivated by the fact that all languages do not provide immediate evidence for every postulated feature. For example, we do not see evidence for honorific features that are clearly represented morphologically in Japanese in Indo-European languages such as English and Polish. Another example can be the existence of the dual number in languages such as Arabic and Slovene, and its absence in, for example, Spanish, French and Italian. The lack of overt realizations of all features cross-linguistically leads to the claim that all features do not have to projected in all languages (cf. Baunaz and Lander 2018c, Haegeman 2003, Rizzi 1994, Shlonsky 2010). According to this interpretation of the universality of the fseq, languages may differ with respect to the number of projected heads, while the relative order of projections has to be preserved. This means that for a set of features [A, B ,C], one language may project them all, while in another language the projected sequence will be shorter, for instance [A, B]. Note again that even if features may be missing from the fseq, the relative order of projections in all languages is considered to be constant.

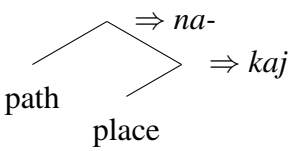
In this analysis I will assume the strict approach, which means that I take the universal sequence to be cross-linguistically uniform. However, regardless of the final answer to the question of the nature of the fseq, the idea of a universal functional sequence has significant consequences for the understanding of linguistic variation in the nanosyntactic model.[38] Since the order in which features are derived in syntax is cross-linguistically invariable, differences between languages do not stem from structure (syntax) but instead depend on how particular languages realize the universal sequence (Starke 2011, 2014). Consider the following representations from Pantcheva (2011), which show goal expressions in Macedonian and Dutch.[39]

[38]Here, I understand linguistic variation as differences in the lexical realization of syntactic structure (Starke 2011).

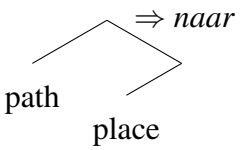
[39]For other nanosyntactic works discussing spatial adpositions see Takamine 2007 and Pantcheva 2008.

Pantcheva (2011) argues that directional phrases can be decomposed into a sequence of two elements, namely the categories of Place (location) and Path (movement):

(118) Macedonian



(119) Dutch



The trees above show the difference between Dutch and Macedonian with respect to the overt realization of Place and Path elements. In Macedonian, the expression corresponding to English *towards* consists of two prepositional morphemes, i.e. the locative *na* and the directional *kaj*. The first morpheme (*na*) realizes only the Path element, while *kaj* is the overt realization of Place. In contrast, Dutch realizes both elements in the the sequence as a single morpheme *naar*. Thus, linguistic variation becomes a matter of the distribution of features inside lexical items.

Lastly, the nanosyntactic model is also based on a few assumptions concerning structure, derivation and structure representation. First, the requirement that all derived structures have to be strictly binary. This standard assumption in generative syntax can be traced back to Kayne (1983) and guarantees unambiguity of structure for the purpose of syntactic mechanisms such as c-command. The second important rule that is universally followed in nanosyntax is that only leftwards movement of phrases (XPs) is permitted. Whenever movement is triggered, it is assumed that the moved element should be adjoined at the top of the structure and thereby extend the root of the phrase marker.[40] The resulting structure will subsequently be linearized in accordance with a structure linearization theory known as the Linear Correspondence Axiom (Kayne 1994):

[40]Compare with the Extension Condition (Chomsky 1993). According to the Extension Condition, syntactic operations have to extend the phrase marker at the root.

(120) Linear Correspondence Axiom (Kayne 1994)

If a non-terminal X asymmetrically c-commands a non-terminal Y, then all terminal nodes dominated by X will precede all terminal nodes dominated by Y.

The asymmetric c-command mechanism that is essential for the LCA is based on the distinction between categories and segments (two directly connected nodes that bear the same label). Asymmetric c-command is defined the following way:

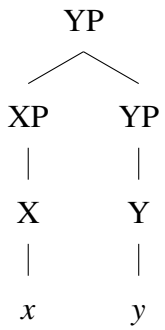
(121) Asymmetric c-command (Kayne 1994: 18)

X c-commands Y iff:

a. X and Y are categories and
b. no segment of X dominates Y and
c. every category that dominates X dominates Y

Therefore, in a representation such as (122), the LCA dictates that *x* has to precede *y*:

(122) Application of the LCA



However, this system cannot be applied in the nanosyntactic framework due to the fact that according to the nanosyntactic system of spellout, phonological exponents are not inserted into terminal nodes but phrasal nodes (XPs). What this means is that linearization has to target phrase nodes instead of terminals (as in the LCA). To solve this problem, Pantcheva (2011) proposes the following reformulation of the LCA:

(123) LCA and phrasal spell-out (Pantcheva 2011: 135)

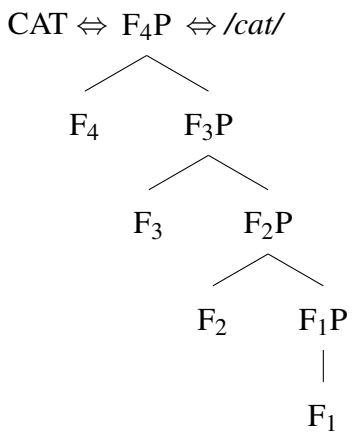
If a non-terminal X asymmetrically c-commands a non-terminal Y, then whatever spells out X precedes whatever spells out Y.

If we apply this version of the LCA to the example in (122), the outcome is that XP has to precede the lower segment of YP.

4.2. The lexicon

The nanosyntactic take on the lexicon stems directly from the idea that each syntactic head equals only one feature (the OFOH maxim). In consequence of this assumption, features are not unordered bundles inside terminal nodes but become ordered syntactic constituents. This means is that what each word or morpheme (a phonological exponent) hides inside is a piece of syntactic structure (an ordered set of features). Given this, Starke (2014: 1) proposes that the lexicon contains only well-formed syntactic expressions. As argued, lexical entries can be reduced to pieces of syntactic structure combined with phonological forms and conceptual information. In consequence, the nanosyntctic lexicon becomes a simple list of well-formed syntactic constituents and their corresponding phonological/semantic information. This way, nanosyntax closely links syntactic structure (the syntactic computational system), phonology (perception and articulation) and pragmatic/conceptual information (lexical semantics). Consider the following example lexical entry in which a piece of syntactic structure [F_1, F_2, F_3, F_4] corresponds to the phonological exponent /cat/ and the conceptual information representing a cat (CAT):

(124) Lexical entry



If the syntactic system generates the structure seen in the example (124), then */cat/* can be inserted into that structure as a matching phonological exponent. This brings us to how lexicalization (spellout) of syntactic structures is accounted for in nanosyntax. The exact mechanism

is however slightly more complex than what the description above might suggest.

In nanosyntax, spellout is understood as a cyclic operation which targets only phrasal nodes (Starke 2009).[41] The phrasal nature of the nanosyntactic spellout stems directly from the *one feature - one head* assumption. As already mentioned, in consequence of this assumption, unordered bundles of features inside lexical items become ordered sets of terminals, which are nothing else than syntactic constituents. Hence, lexical items are phonological forms inserted into pieces of syntactic structure (XPs). Lexical insertion (spellout) is also cyclic, since it is triggered whenever a new XP is formed, that is with every feature merge. Each time a new feature is added to a derivation, the spellout mechanism checks the lexicon in search for a matching entry (a piece of structure stored in the lexicon). If a matching entry is found, then the phonological exponent corresponding to that lexical entry can be inserted into the derived structure. Below, I will discuss an example in which the syntactic system has to derive the sequence [F_1, F_2, F_3]. The following four lexical entries are available. I will use these entries throughout the rest of the chapter:

(125) Lexical entries

a. /α/ ⇔ F_1P
|
F_1

b. /β/ ⇔ F_3P
/ \
F_3 F_2P
/ \
F_2 F_1P
|
F_1

c. /λ/ ⇔ F_3P
|
F_3

[41] See also Baunaz and Lander 2018c, Caha 2020, Starke 2014, 2018.

d. /γ/ ⇔ F_4P

F_4

e. /δ/ ⇔ F_5P

F_5 F_4P

F_4 F_x

With the merger of the first feature [F_1], the lexicalization system accesses the lexicon and finds a lexical entry matching the derived phrase, namely (125-a)). Having successfully obtained a match, the spellout system will insert the phonological exponent corresponding to the matching lexical entry (α) into the derived phrase:

(126) Lexicalization of F_1P

/α/ ⇐ F_1P

F_1

The next merge-spellout cycle begins with the addition of F_2, which is the next feature in the specified feature sequence. However, this time, the lexicon does not contain a lexical entry that can become an exact match for the derived structure. This poses a problem, due to the fact that features may not be left out or removed from the computation component. All features provided by the fseq have to undergo spellout and be properly realized as lexical items before the next merge cycle may begin.[42] To prevent derivations from crashing when an exact match is not found in the lexicon, lexical insertion in the nanosyntactic model is regulated by the Superset Principle (Starke 2009: 3) which states that a tree corresponding to a lexical entry matches a particular syntactic structure if it constitutes a superset of that structure. In other words, a structure may be spelled out with a phonological exponent corresponding to an overspecified lexical entry. The Superset principle may be formulated the following way:[43]

[42]This is known as the Exhaustive Lexicalization Principle (Fábregas 2007).

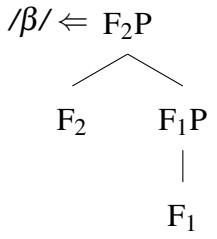
[43]Note that according to the Superset Principle, a lexically stored tree matches a syntactic tree (a node) that it is

(127) The Superset Principle (Baunaz and Lander 2018c: 39)

"A lexically stored tree L matches a syntactic tree S if L is a superset (proper or not) of S."[44]

This means that entry (125-b) may become a match for [F_1, F_2] (since it constitutes a superset of this structure). The insertion of a new phonological exponent (β) into the derived constituent will also overwrite the previously inserted exponent (α). In other words, the phonological exponent inserted in the previous spellout cycle will be replaced by the phonological exponent inserted in the current spellout cycle. The resulting representation is as follows:

(128) Lexicalization of F_2P

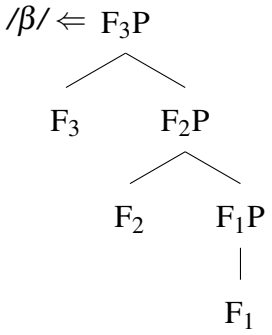


The merger of the last feature in the sequence (F_3) again forms a structure that can be lexicalized with one of the available lexical entries. This time, entry (125-b) will become an exact match for the derived structure. As in the previous cycle, the new phonological exponent (β) will also overwrite the outcome of the previous spellout (also β in this case). The three merge-spellout cycles will generate the following structure spelled out as β:

a superset of, not a set of features that constitute a subset of the lexically stored tree. For this reason, a lexically stored structure such as [F_3[F_2[F_1]]] will match [F_2[F_1]] but not [F_3[F_1]]. The lexically stored tree [F_2[F_1]] does not constitute a superset of the tree [F_3[F_1]].

[44] A proper superset of set X is a superset of X that is not equal to X.

(129) Lexicalization of F_3P



The Superset Principle guarantees that a structure may be spelled out even if there is no entry in the lexicon of a language that would be a perfect match. The derived structure may still be lexicalized with an overspecified lexical entry, that is a larger entry which contains the derived structure as its subset. Also note that the Superset Principle means that a single lexically stored tree may be used to lexicalize multiple syntactic structures as long as those structures are contained within one another. In the derivation describe above, the lexical entry in (125-b) is used to lexicalize [F_1, F_2] as well as [F_1, F_2, F_3].

The Superset Principle is, however, not the only rule that governs lexical insertion in nanosyntax. In (126), when the first feature of the sequence is merged (F_1), entry (125-b) could also be considered a match for the derived structure as it constitutes its superset. The spellout system will however always choose (125-a)) over (125-b) because of the rule known as the Elsewhere Principle. In accordance with this principle, the spellout system will always choose a lexical entry with the smallest number of superfluous features. In other words, the system always picks the best match:

(130) The Elsewhere Principle (Baunaz and Lander 2018c: 44, Starke 2009: 4)
"If several lexical items match a syntactic node, insert the entry with the fewest features unspecified for that node."

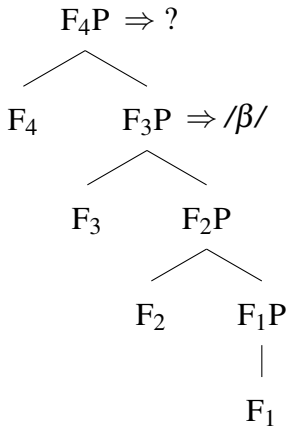
The Superset Principle and The Elsewhere Principle constitute two very closely connected aspects of the nanosyntactic spellout system. While the Superset Principle makes it possible to lexicalize multiple structures with a single lexical entry, the Elsewhere Principle solves the problem of lexical entry competition by ruling out the possibility that an entry will become a

match if there is any lexical entry that constitutes a better match for the derived structure.[45]

4.3. Spellout-driven movement

The aspects of the nanosyntactic theory discussed so far do not account for a rather important theoretical issue, that is the possibility that the lexicon may not contain any matching lexical entries. This means cases in which a structure cannot be spelled out with any lexical entry available in the lexicon, even under the Superset principle. In nanosyntax, the solution to this problem is also the answer to the question of how languages may spell out the universal sequence in different ways. Whenever the spellout system cannot locate a matching lexical entry in the lexicon, it will try to save the derivation from crashing by triggering syntactic movement. In other words, movement in nanosyntax may be caused by the need to obtain a lexicalizable tree geometry. This is known as the spellout-driven movement mechanism (Caha 2020: 25-33, Starke 2018). Consider the following situation in which F_4 has just been added to the previously derived structure:

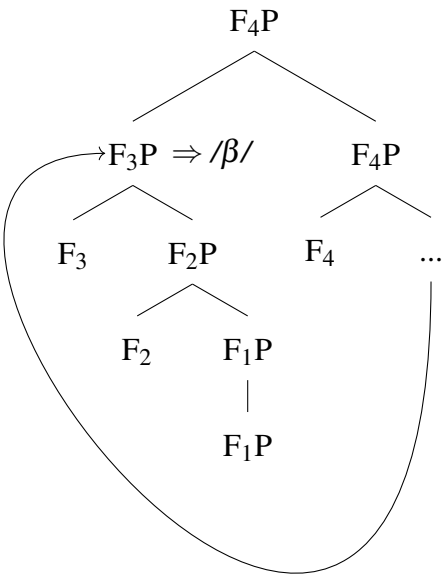
(131)



Since there are no lexical entries that could become a match for the derived structure, spellout-driven movement has to trigger as a last-resort operation. In order to obtain a structure that can be spelled out with one of the available lexical entries, F_2P has to be displaced (rolled-up) to a position above F_4, which will result in a tree where F_4P is a constituent containing only F_4. The remerged phrase will become similar to what more traditional theories consider a specificer; it will extend the root of the tree, but it will nor project. Additionally, it is assumed that movement triggered for the purpose of spellout does not leave a trace:

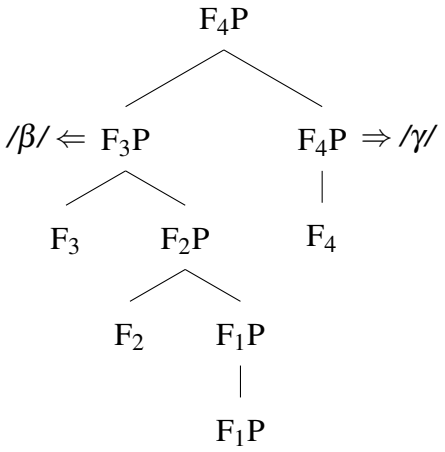
[45] See also Caha 2009: 55.

(132) Spellout-driven movement



In consequence of the roll-up (snowball) movement, F_4P will be stranded and become a constituent matching (125-d). A successful match means that F_4P will be lexicalized as γ and become a morpheme separate from the output of the previous lexicalization cycle (β) (a suffix):

(133)



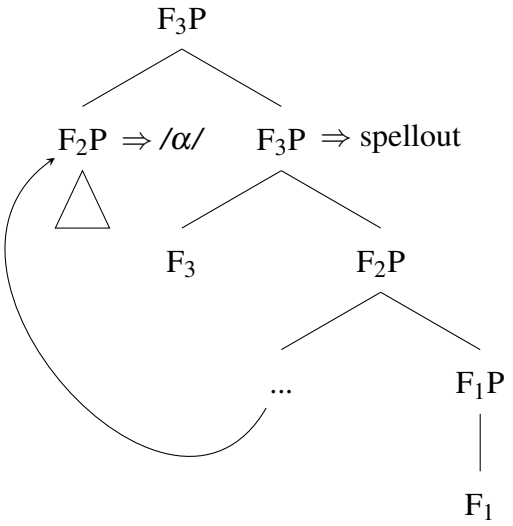
As proposed in Starke (2018), the spellout-driven movement mechanism always follows a specific algorithm. Unsuccessful spellout will always be followed by cyclic (spec-to-spec) movement and then roll-up (snowball) movement:

(134) Spellout algorithm (Starke 2018: 357)

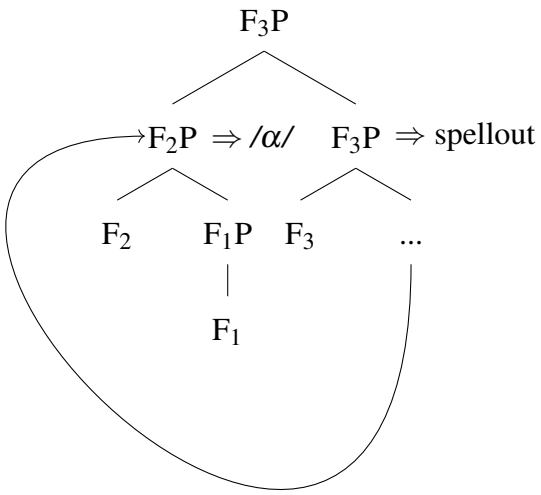
a. insert feature and spell out
b. if fail, try a cyclic (spec-to-spec) movement of the node inserted at the previous cycle and spell out
c. if fail, try a snowball movement of the complement of the newly inserted feature and spell out

This means that the following two scenarios may occur if a suitable lexical entry is not found after f-merge. First, the system will attempt to displace an XP inserted as a left branch in the previous cycle (spec-to-spec movement). If this fails to produce a lexicalizable tree geometry (or cannot be performed due to a lack of an XP), a roll-up (snowball) movement will occur next leading to the displacement of the complement. Compare the following two representations:

(135) Spec-to-spec movement



(136) Roll-up (snowball) movement

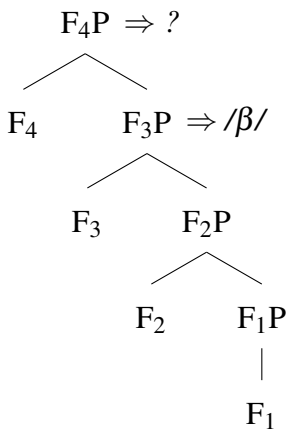


However, there still exists a possibility that none of the steps above will produce a lexicalizable structure. Should this happen, the system will have to consider the last merge-f operation incorrect and undo it. Subsequently, the next step of the spellout algorithm will be applied at the previous merge cycle (cyclic and then roll-up movement):

(137) Backtracking (Starke 2018: 358)
If the spellout algorithm fails, go back to the previous cycle, and try the next option for that cycle.

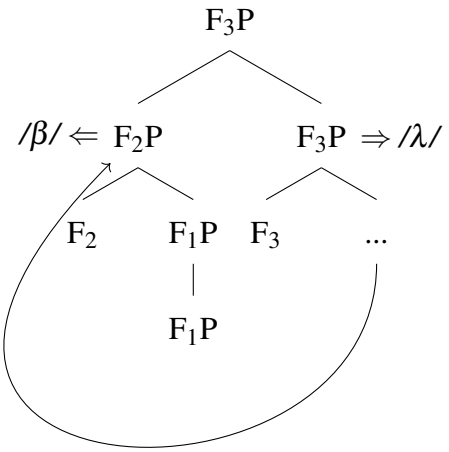
This step in the algorithm is known as backtracking, and if necessary, it may be attempted multiple times until a lexicalizable tree geometry is obtained. Of course, each backtracking operation has to be followed by transformations, that is spec-to-spec and roll-up movements, before another f-merge may be undone. Consider the following scenario in which F_4 is the next feature in the fseq and the outcome of the previous spellout cycle was β (lexical entry (125-b)):

(138)



The lack of a matching lexical entry will trigger spellout-driven movement which will result in the displacement of the complement of F_4 (F_3P). However, assume that ,unlike in the previous example, the lexical entry in (125-d) is not available and the stranded F_4P cannot be spelled out at γ. Due to the fact that spellout-driven movement has failed to form a tree in which F_4 can be lexicalized, the spellout system will consider the merger of F_4 incorrect and initiate backtracking. In consequence F_4 will be removed from the derivation and spellout driven movement will be applied at F_3. In the resulting tree, F_3P will become a stranded constituent for which (125-c) is a match. Consequently, F_3P will be lexicalized as λ:[46]

(139)



If backtracking fails, even after multiple attempts, there is still one last option left. The last last-resort operation that triggers whenever everything else has failed (movement and backtracking) is merge-XP. This operation involves the formation of a specifier-like XP through a

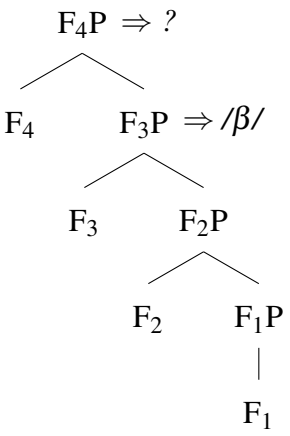
[46]Note that F_2P will receive the exponent inserted before F_3 was merged, that is β (lexical entry (125-b)).

separate subderivation. The subderived phrase will contain the desired feature, and when it is spelled out, it will be integrated into the main structure. Due to its complex nature (generate and merge XP), the process of subderivation is considered much more costly than merge-f and remerge (movement) and therefore placed as the final last-resort step in the spellout algorithm:

(140) Subderivation (Starke 2018: 359)
"If merge-f has failed to spellout (even after backtracking), try to spawn a new derivation providing feature X and merge that with the current derivation; projecting feature X to the top node."

Consider the following example where again the fseq dictates that F_4 is the next feature to be merged. The output of the previous lexicalization cycle was β (lexical entry (125-b)):

(141)



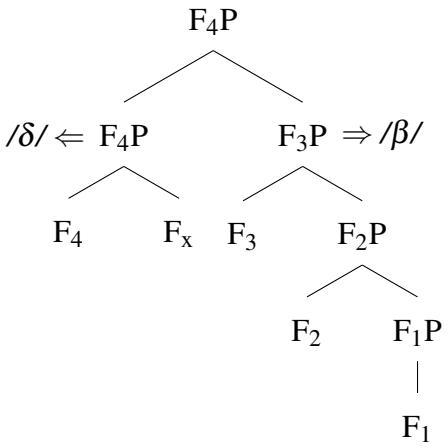
If there is no lexical entry that can match the derived sequence, the spellout system will attempt to apply the first two last-resort operations (cyclic and roll-up movements). Now assume that the lexical entries in (125-d) and (125-c) are not available. This means that spellout-driven movement (roll-up) will not create a lexicalizable tree geometry. Once the first attempt to save the derivation through movement has failed, the system will undo the merger of F_3 and go back to the previous cycle to apply transformations at that point. However, in this scenario, movement after backtracking will also not result in a tree that can be fully lexicalized.[47]

Only after all other options have not provided a structure that can spell out F_4, will the spellout system attempt to subderive an XP containing this feature. The subderivation will begin with a merge operation in order to form a minimal phrase containing F_4. Since there is no clear consensus regarding the binary bottom of a subderivation, I assume that the system

[47] As mentioned, backtracking will occur multiple times if a lexicalizable tree geometry cannot be found.

will provide whatever base feature is necessary to spell out F_4 (here labeled F_x).[48] The lexical entry in (125-e) will be used to lexicalized the subderived constituent. Once spelled out, the subderived phrase can be integrated into the main structure with F_4 projecting to the top node. This will create a morpheme that will be linearized on the left side of β:[49]

(142) Subderivation: prefix formation



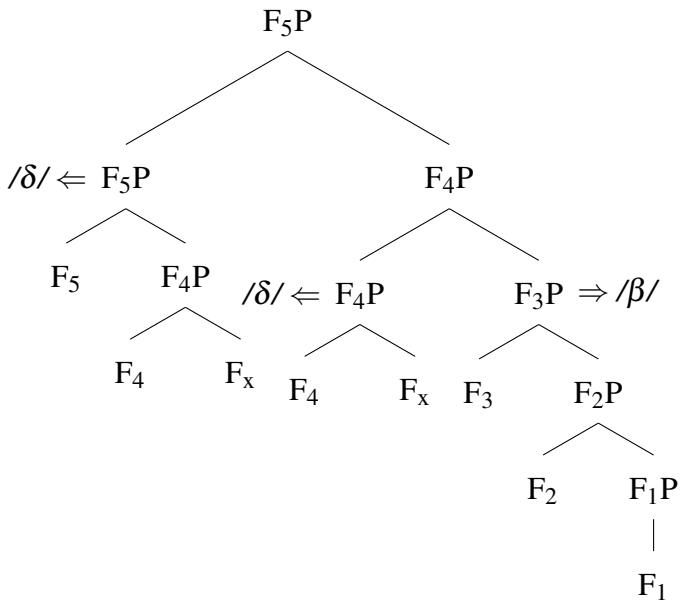
The steps shown above allow Nanosyntax to derive specifier-like complex left branches (XPs), which, unlike outcomes of spellout-driven movement, will become morphemes surfacing on the left side of the stem (prefixes). However, this system, as it has been presented so far, makes it possible to construct only very simple phrases and faces a certain problem.

When at the next merge cycle F_5 is added to the structure, the only way to spell it out will be to generate a subderivation providing this feature. If we follow the algorithm steps presented above, we will arrive at the following suboptimal result where another subderivation had to be spawned to lexicalize F_5 (using the lexical entry in (125-e)):

[48] Caha et al. (2019) suggests that a subderivation begins with a copy of the last successfully merged feature from the main spine. While this is possible, I will however not explore this idea here.

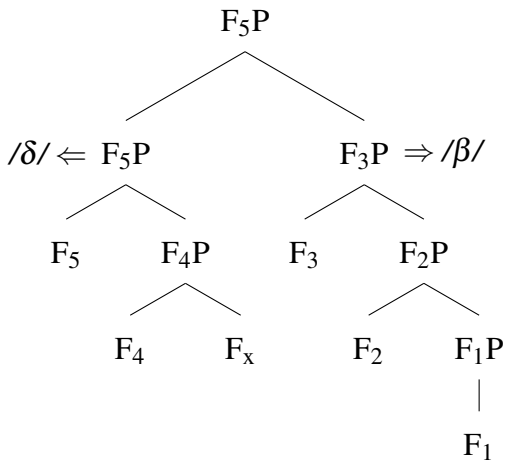
[49] It should be noted that the fact that prefixes are assumed to project constitutes an issue from the perspective of LCA. However, this issue can be solved if we take into account the fact that that subderivation involves merge-XP. The topmost node projected in consequence of spellout-driven movement is the extension (segment) of the right-branch node dominating the stranded constituent. The displaced constituent does not project to the topmost node. In contrast, subderivation means that the main derivation is merged with a subderived XP. Thus, the projected node is projected from the combination of both phrases and is not a segment of the subderived XP. In consequence, the subderived phrase will asymmetrically c-command the main derivation and precede its exponent in linear order.

(143) Suboptimal prefix spellout



To prevent the spellout algorithm from spawning a new subderivation each time a feature can be lexicalized only as part of a complex left branch, Starke (2018) proposes that a subderivation may remain active as long as it is necessary. Since subderive may be considered the most costly operation for the syntactic systems, a subderivation will be kept open in order to merge and lexicalize as many features as possible. Only when the subderivation can provide and spell out no more features, will it be integrated into the main structure, as shown in the following representation:

(144)



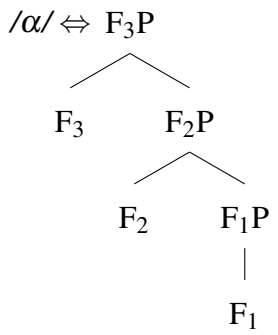
When F_5 is to be merged as the next feature in the fseq, instead of spawning another subderivation, the system will simply ask the current subderivation to provide another feature through

regular f-merge (the regular rules of the spellout algorithm will apply). Only when the subderivation can no longer be used to provide and lexicalize features, can it be closed and added to the main derivation.

4.4. Nanosyntax: consequences

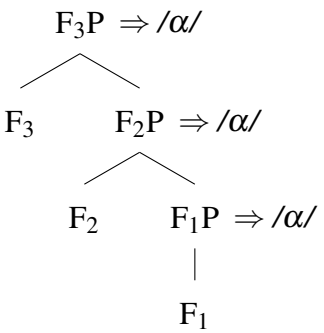
The nanosyntactic model of derivation successfully accounts for the phenomenon of syncretism and the *ABA generalization. Whenever syncretism arises in a paradigm, it is the consequence of the Superset Principle, which allows the spellout system to lexicalize multiple structures with a single lexical entry. Of course, the lexicalized pieces of syntactic structure have to exist in a containment relation, which means that smaller structures have to be syntactically contained within larger ones. Consider one more example which shows the application of the Superset Principle. To spell out a sequence of heads such as [F_1, F_2, F_3], the following lexical entry may be used:

(145) Lexical entry



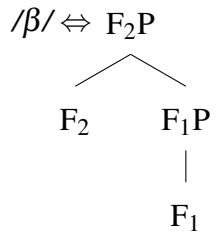
The lexical entry in (145) will be used to spell out not only the whole sequence once F_3 has been merged but also all its subsets at earlier stages of the derivation. In consequence, the three structures that can be derived on the basis of the sequence of [F_1, F_2, F_3], that is [F_1], [F_2[F_1]] and [F_3[F_2[F_1]]], receive the same phonological exponent:

(146)



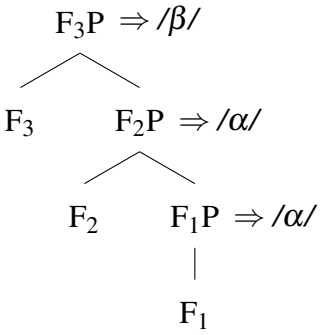
As for the *ABA generalization, the ABA pattern of syncretism is not possible in the nanosyntactic system of spelllout due to the Elsewhere Principle. Assume that apart from the entry in (145), another entry is available for the derivation of the sequence of [F_1, F_2, F_3]:

(147) Lexical entry



After the first f-merge, both (147) and (145) can become a match for the structure ([F_1]). However, because of the Elsewhere Principle, the spellout system will never choose the entry with a larger number of superfluous features ((145)). The lexical entry in (147) has to become the match in the first two spellout cycles. The entry in (145) will become a match only in the last cycle since (147) does not contain F_3 and therefore cannot be used to spell out the whole sequence. The Elsewhere Principle rules out the possibility of the ABA pattern emerging as the outcome of cyclic override of one phonological exponent by another. The correct output of the discussed derivation will be (148):

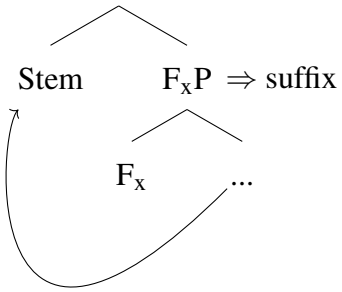
(148) Correct spellout outcome



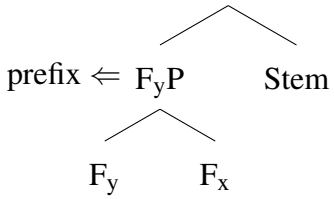
Another interesting consequence of the nanosyntactic derivation system and spellout is that there is now a very specific difference between morphemes emerging on the right (suffixes) and on the left (prefixes) of a stem. This difference lies not just in the surface placement of such morphemes but directly stems from the two distinct ways of forming new constituents in nanosyntax, that is spellout-driven movement and subderivation. This ties in nicely with other aspects of the theory that reduce morphology and syntax into a single computational and structure-forming component of language (see Section 4.1.):

(149) Suffix vs. prefix

a. Suffix



b. Prefix



The morphosyntactic system creates a suffix through evacuating a piece of structure (snowball movement), which results in the formation of a remnant constituent containing a unary grouping at the bottom (since the displaced constituent does not leave a trace). In contrast, prefixes are XPs formed through subderivation. As spawning a separate derivation space always begins with a merger of two features, prefixes will always have a binary grouping at the bottom.

Chapter 5: Analysis

5.1. Indefinite markers: the proposal

The analysis below presents the application of the principles of nanosyntax for the derivation of non-specific, specific unknown and specific known indefinite markers. Each of the attested patterns of syncretism is discussed separately and through relevant examples illustrating the particular steps necessary for the derivation of that pattern. As already mentioned, the emergence of the patterns of syncretism is connected with the number of available lexical entries that can be used to spell out the proposed indefinite hierarchy. Depending on the number of lexical entries, we will see two or three subsets of the structure being lexicalized with a single lexical entry. Furthermore, the analysis also accounts for the possible positions of the marker with respect to the stem, i.e. a prefix or suffix, and the attested combinations of prefixal and suffixal indefinite markers in the paradigms of languages in which two or three phonological exponents are used to realize the indefinite structure.

The analysis of the collected indefinite pronoun data shows that if arranged in accordance with the order shown in the implicational map of indefinite functions proposed in Haspelmath (1997), indefinite markers corresponding to the non-specific, specific unknown and specific known indefinite functions reveal a number of patterns of syncretism (AAA, ABB, AAB). At the same time, the ABA pattern remains unattested, as predicted by the *ABA generalization (Bobaljik 2007, 2012). Table 9 shows the patterns of syncretism attested in the studied language sample:

Table 9. Indefinite markers: patterns of syncretism.

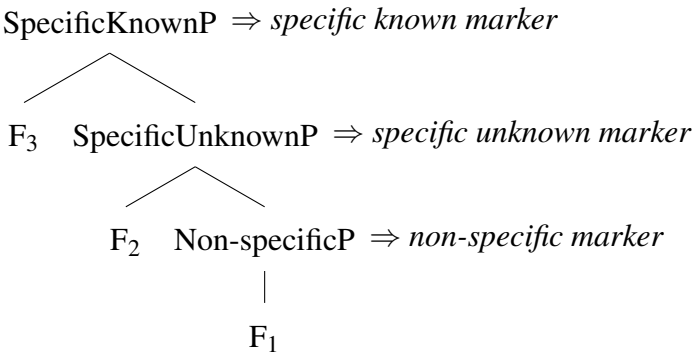
language	non-specific	specific unknown	specific known	pattern
English	some-	some-	some-	AAA
Georgian	-me	-γac	-γac	ABB
Latin	ali-	ali-	-dam	AAB
Unattested				ABA

The observed patterns and the lack of the ABA pattern constitute strong evidence that non-specific, specific unknown and specific known indefinite markers form a paradigm in which the relative order of elements conforms to the order shown in the map of indefinite functions (Haspelmath 1997).

As shown in Chapter 4, we can use the methodological tools provided by the nanosyntactic framework to successfully explain the emerging patterns of syncretism as well as the absence of the ABA pattern. The nanosyntactic model of grammar predicts that syncretism will stem from the application of the Superset Principle which allows a single lexical entry to spell out multiple syntactic structures as long as these structures remain in a containment relation and the features they consist of form a hierarchy (a sequence). Additionally, the Superset Principle is constrained by the Elsewhere Principle which guarantees that the spellout system always selects the lexical entry which constitutes the best match. In consequence of this principle, we should not observe the ABA pattern in ordered sets of forms.

Taking into account the results of the cross-linguistic analysis of non-specific, specific unknown and specific known indefinite markers and the nanosyntactic view of syntactic derivation mechanisms and spellout, we can propose the non-specific, specific unknown and specific known indefinite functions are derived on the basis of a hierarchy of features that consists of three layers of structure (Dekier 2021). Each marker type, that is a marker representing a particular indefinite function, will contain a different set of features from the hierarchy as a result of feature stacking. This kind of cumulative representation means that the three types of markers, understood as particular structures, constitute three different subsets of the hierarchy and are syntactically contained within one another. In consequence, the three marker types may be lexicalized syncretically under the Superset Principle:

(150) Indefinite hierarchy



As mentioned above, the attested patterns of syncretism result from multiple layers of the indefinite hierarchy being lexicalized with the phonological exponent of a single lexical entry. Particular languages differ with respect to the number of lexical entries that are used to spell out the indefinite hierarchy which means different patterns of syncretism.

However, syncretism alone shows only the relative order of elements in a syntactic hierarchy, which means that it does not indicate which structure is the simplest, and which should be considered the most complex. For this reason, it is necessary to find a criterion that can help us establish the least and the most complex of the three indefinite marker types. A criterion that I will use to provide a solution to the issue is the semantic compositionality of the indefinite markers in question. As the semantics of the three indefinite markers indicate, the non-specific indefinite markers should be considered the simplest of the three types since they introduce only a referent variable (category) without narrowing the referent down to a specific individual. Specific unknown markers are in the middle of the hierarchy and add specificity to the referent. Lastly, specific known markers can be considered the most complex of the three types, due to the fact that they convey the knowledge of the speaker on top of referent specificity.[50] The referent has to be specific first if it is to be known by the speaker, which clearly suggests that specific known markers are higher in the hierarchy than specific unknown marker (and non-specific markers for that matter). This reasoning is the consequence of the assumption that as the syntactic hierarchy grows, so should the semantic complexity of the derived markers. If the assumption correctly predicts the relationship between structure and semantics, then we are compelled to claim that (150) shows the correct order of features in the indefinite hierarchy.

[50]I use the terms *specificity* and *knowledge of the speaker* to highlight the differences between the three types of indefinite pronouns. I do not intend to discuss these terms as representing particular grammatical properties.

The proposed indefinite hierarchy can be used as the basis for the derivation of non-specific, specific unknown and specific known indefinite markers. The analysis presented in Chapter 5 will explain from the perspective of syntactic derivation not only the emergence of the patterns of syncretism attested crosslinguistically but also account for the unavailability of the ABA pattern. Additionally, the syntactic models put forward in the analysis will shed some light on the morphological positioning of indefinite markers with respect to base stems. This means that the nanosyntactic derivation system will allow us to propose representations for prefixal and suffixal markers, as well as the change of one type of marker into another within a paradigm. However, before we can take a closer look at the morphosyntax of non-specific, specific unknown and specific known indefinite markers, it is necessary to devote some space to a more detailed discussion of indefinite pronoun bases. In the next section, I will analyze the structure of the available indefinite pronoun bases and how they can be connected with the proposed indefinite hierarchy.

5.2. The base

The aim of this section is to present the attested forms of the indefinite pronoun base. We can distinguish its main types on the basis of their form and structure, namely generic noun bases, pronominal bases and bases derived from the numeral *one*. They are illustrated with the following examples, in which the base is in bold:

(151) Noun-based indefinites

a. *some-**thing***

b. *någon- **ting***
INDEF thing
'something' (Swedish)

c. *bir- **şey***
INDEF thing
'something' (Turkish)

(152) Pronoun-based indefinites

a. ***co*** *-ś*
what INDEF
'something' (Polish)

b. ***vala*** *-mi*
INDEF what
'something' (Hungarian)

c. ***nani*** *-ka*
what INDEF
'something' (Japanese)

(153) Generic-*one*-based indefinites

a. *some-**one***

b. *qualc-* ***uno***
INDEF one
'someone' (Italian)

c. *quelqu'-* ***un***
INDEF one
'someone' (French)

All three types of forms may be used to represent different categories of referents. This is most clear with noun bases as their lexical meaning is usually very close to the category they represent:

(154) Noun-based indefinites

a. *qual-* ***cosa***
INDEF thing
'something' (Italian)

b. *bir-* ***yerde***
INDEF place
'somewhere' (Turkish)

c. *xi-* ***darba***

INDEF time

'sometime' (Maltese)

Nominal bases may be used to form multiple pronouns in a series. For example, in the Maltese *xi-* series, almost all bases constitute stand-alone generic nouns. The exceptions are *ħadd* in *xi-ħadd* 'somebody', which comes from the Classical Arabic *aħad* 'one', and *mkien* in *xi-mkien*, which is related to the word *makaan* 'place' (Haspelmath 1997: 296-297):

(155) Maltese (Haspelmath 1997: 296-297)

a. *xi-* ***ħadd***

INDEF PERSON

'someone'

b. *xi-* ***ħaga***

INDEF thing

'something'

c. *xi-* ***mkien***

INDEF PLACE

'somewhere'

d. *xi-* ***darba***

INDEF time

'sometime'

e. *b'xi-* ***mod***

INDEF manner

'somehow'

A similar paradigm is the Turkish *bir-* 'some-' series, in which only the form corresponding to English *somebody* does not have a base that can be a stand-alone noun (Haspelmath 1997: 286-287):[51]

[51]The form *biri* is derived from *bir* 'one'.

(156) Turkish

a. *biri(si)*
one
'someone'

b. *bir-* ***şey***
INDEF thing
'something'

c. *bir-* ***yer****-de*
INDEF place-LOC
'somewhere'

d. *bir-* ***zaman***
INDEF time
'sometime'

e. *bir-* ***şekil****-de*
INDEF manner-LOC
'somehow'

Another category of bases consists of two types of forms, namely interrogative pronouns and relative pronouns. I list these two kinds of forms together as pronominal bases due to their similarity or even surface identity in some languages. The examples in (157) show how in English the same lexical item 'where' can be used as an interrogative pronoun, a relative pronoun and an indefinite pronoun base:[52]

(157) a. ***Where*** *did you go after the party?* interrogative pronoun
b. *The city* ***where*** *he lives is beautiful.* relative pronoun
c. *some-****where*** indefinite pronoun base

However, the word *how*, which appears as a base in *some-how*, can only be used as an interrogative pronoun:

[52] See also Baunaz and Lander 2018b, where interrogative pronouns and relative markers are argued to be derived on the basis of the same hierarchy of features.

(158) a. ***How*** *did you do that?* interrogative pronoun

b. **The way* ***how*** *he looks at her is disturbing.* relative pronoun

c. *some-****how*** indefinite pronoun base

The fact that English *how*, which is not used as relative pronouns, appears as indefinite pornoun base suggests that English employs interrogative pronouns as bases, rather than relative pronouns. The use of interrogative pronouns as bases is also clear in Japanese which lacks relative pronouns:

(159) Japanese (Matsumoto 1997: 43)

Donarudo Toranpu-ga katta mise wa doko?
Donald Trump-NOM bought store TOP where
'Where is the store (which) Donald Trump bought?'

The assertive indefinite pronoun series in Japanese is therefore based on interrogative pronouns. Example (160) provides a list of assertive indefinites in Japanese:

(160) Japanese

a. ***dare*** *-ka*
who INDEF
'someone'

b. ***nani*** *-ka*
what INDEF
'something'

c. ***doko*** *-ka*
where INDEF
'somewhere'

d. ***itsu*** *-ka*
when INDEF
'sometime/some day'

e. ***dō*** *-ka*

how INDEF

'somehow'

Relative pronouns are also attested as bases of indefinite pronouns. For instance, Finnish makes use of its relative pronoun *jo-ka* 'which/who/that' to form its assertive indefinite series. As shown in the example below, the relative pronoun *joka* cannot be used as an interrogative word (*mikä* 'what' and *kuka* 'who' have to be used):

(161) Finnish Karlsson (1983)

a. *Hän on mies*, ***joka*** *ei pelkää.*

he is man REL not fears
He is a man **who** does not fear.

b. ***Mikä/*joka*** *tämä on*?

what this is
What is this?

c. ***Kuka/*joka*** *tuo pitkä nainen on*?

who that tall woman is
Who is that tall woman?

The assertive indefinite pronoun series in Finnish, which is shown below, bears the textit-kin (with the exception of *jo-ku* 'someone'). The different referent categories in (162-c), (162-d) and (162-e) are encoded through case suffixes (in bold) added to the relative stem *jo-* (exceptions are *jo-ku* 'someone' and *jo-kin* 'something').

(162) Finnish (Haspelmath 1997: 293)

a. *jo -ku*

REL INDEF

'someone'

b. *jo -kin*

REL INDEF

'something'

c. *jo* ***-ssa*** *-kin*

REL INESS INDEF

'somewhere' (inessive case)

d. *jo* ***-lloin*** *-kin*

REL TEMP INDEF

'sometime' (temporal case)

e. *jo* ***-ten*** *-kin*

REL CAUS INDEF

'somehow' (causative case)

Lastly, I provide examples showing generic *one*-pronoun bases. Indefinite pronouns formed on the basis on generic *one*-pronouns can be found in, for example, English, French and Italian:

(163) a. *some-* ***one***

b. *qualc-* ***uno***

INDEF one

'someone'

c. *quel'qu-* ***un***

INDEF one

'someone'

I discuss the use of one-pronoun in English, French and Italian in Section 5.3.

5.3. Ontological categories

The referent categories expressed by bases can be classified as elements of a larger group known as ontological categories. Some of such categories which have been recognized in the literature are THING, PERSON, PLACE, TIME, MANNER and AMOUNT (Baunaz and Lander 2018a, Cinque 2008, Haspelmath 1997: 21-31, Jackendoff 1983: 51).[53] In the following sections, I will argue that ontological categories are syntactically represented as functional nominals inside nominal, pronominal and generic *one*-pronoun bases used to form indefinite pronouns.

[53] I will use capital letters for ontological categories.

The existence of functional nouns is discussed as early as in Bliese (1981: 15-30). As shown in example (164), the Afar laguage (Cushitic) uses the functional noun *gi'de* in relative constructions to express meanings such as *how much* (amount) and *how long* (time):

(164) Afar

a. *A'nu 'usuk y-a-k'me-h* ***gi'de*** *aadi'ge*.
I he he-IMPF-eats-h amount I.know
'I know the amount he eats.' (Bliese 1981: 26)

b. *'sin 'cari el'le y-acusbusee-'ni-h* ***gi'de*** *taadi'gee*.
your house until they-will-repair-PL-h how.long you.know
'Do you know how long it takes them to repair your house?' (Parker 1974:42)

In relative constructions corresponding to English sentences with *the place which* and *where*, Afar uses an unspecified noun in combination with the clitic *-ikke* which appears to be related to *ak'ke* 'place':

(165) Bliese (1981: 28)

'oson ***(a)'kah*** *y-inniki'se-n-ikke t-ub'le*.
they NP they-fell-pl-place you.saw
'You saw the place where they fell.'

Another significant step forward in understanding the place of ontological categories in grammar is Kayne (2005), who proposes the presence of silent nouns in nominal constructions. While having null exponents, these functional elements were considered to be cross-linguistically universal. In the following example, the words in brackets represent unpronounced nominals:

(166) Functional nominals (Kayne 2005: 145, 242, 249)

a. *a few [NUMBER] books*
b. *a little [AMOUNT] money*
c. *a green [COLOR] car*
d. *the age of eleven [YEARS]*

A similar analysis is presented in Cinque (2008), where headless relative clauses are claimed to contain silent functional heads.[54] Such heads are argued to form a closed class composed of the following categories: PERSON, THING, PLACE, TIME, AMOUNT and MANNER:

(167) Silent functional heads (Cinque 2008: 18)

a. *(I don't like) [[what THING you said] (SUCH) THING]*
b. *(He weighs) [[what AMOUNT you weigh] (SUCH) AMOUNT]*
c. *(Here is) [[where PLACE they slept] THERE PLACE]*
d. *(I was there) [[when TIME he said that] THEN TIME]*
e. *(She hates [[whoever PERSON does that] (SUCH) PERSON]*

The functional heads proposed in Cinque's analysis are however not always silent and, in some languages, they are overtly realized. This is achieved through the use of functional nominals. A language in which functional nouns are overtly realized is Lango (Nilo-Saharan):

(168) Lango (Noonan 1992: 220)

Marô ***gin*** [*amê* *camô*].
3SG.like.HAB thing REL+PART 3SG.eat.HAB
'He likes what he eats.'

In Lakhota, functional nouns remain in situ within relative clauses and are optional:[55]

(169) Lakhota (Cinque 2008: 18)

[*Mary* (***taku***) *kaǧe*] *ki*] *ophewatų*
Mary (thing) make the I-buy
'I bought what Mary made.'

[54] See also Cinque (2020: 97-101).

[55] In the original example shown in Cinque (2008: 18), *taku* is glossed as *something*. The word *taku* may mean either *thing*, *what* or *something*. In this particular case, it seems justified to treat *taku* as a dummy noun (thing). Below, I provide the original example:

i [Mary (**taku**) kaǧe] ki] ophewatų
M. (something) make the I-buy
I bought what Mary made. (Lakhota, Cinque 2008: 18)

Functional nouns corresponding to other categories such as PERSON and PLACE are also attested. The following examples are from Yidiɲ (Australian) and Bih (Malayo-Polynesian):

(170) a. ***Bama:l*** *yabuɽuɲgu mia* *gangu:l* *wawa:l.*
person-ERG girl-ERG animal-ABS wallaby-ABS see.PAST
'The person girl saw the animal wallaby.' (Yidiɲ, Dixon 1977: 480)

b. *Ti* ***anôk*** *ŏng dôk?*
where place you stay
'Where are you?' (Bih, Nguyen 2013: §6.2.1.5)

The presented examples suggest the existence of a functional nominal category representing what is known as ontological categories, that is concepts such as THING, PERSON, PLACE, MANNER, TIME, AMOUNT, etc. (Baunaz and Lander 2018a). As shown, these categorical nominals are used in order to determine the class of a noun phrase or type of reference in a relative construction (e.g. (170) and (168)).

Functional nouns of the kind shown in the examples above can be argued to be realized by the three types of indefinite pronoun bases. The first type of bases, that is nominal bases, indicates the presence of functional in the most straightforward way. As mentioned in Section 5.2, nominal bases tend to take the form of nouns which lexically closely correspond to the referent categories they represent. Since this is very similar to how functional nominals are shown to be overtly realized in (168), (169) and (170), it can therefore be argued that nominal bases are also overt realizations of functional nouns used for the purpose of determining the category of the referent. In English, the bases *body*, *thing* and *place* will thus realize the categories PERSON, THING and PLACE:

(171) a. *some- body*
INDEF PERSON

b. *some- thing*
INDEF THING

c. *some- place*
INDEF PLACE

Generic nouns representing other ontological categories such as MANNER and TIME are at-

tested in other languages:

(172) a. *xi-* ***darba***
INDEF TIME
Maltese: 'sometime'

b. *b'xi-* ***mod***
INDEF MANNER
Maltese: 'somehow'

c. *bir-* ***şekil****-de*
INDEF MANNER
Turkish: 'somehow'

d. *bir-* ***zaman***
INDEF TIME
Turkish: 'sometime'

Interrogative pronouns can also be shown to contain functional nominals. In the following example from Icelandic, we see that for each referent category a different morpheme is added to an invariable wh-question element:

(173) Icelandic

a. *hv-* ***er***
WH PERSON
'who'

b. *hv-* ***að***
WH THING
'what'

c. *hv-* ***ar***
WH PLACE
'where'

d. *hv-* ***enær***
WH TIME
'when'

e. *hv-* ***ernig***
WH MANNER
'how'

Lithuanian interrogative pronouns can be analyzed the same way. An invariant interrogative element is added to a morpheme representing an ontological category:

(174) Lithuanian

a. *k-* ***as***
WH PERSON/THING
'who/what'

b. *k-* ***ur***
WH PLACE
'where'

c. *k-* ***ada***
WH TIME
'when'

d. *k-* ***aip***
WH MANNER
'how'

In English, *what*, *where* and *when* can be decomposed into a Wh- element and a categorical morpheme, whereas who and how should be analyzed as

(175) English

a. *wh-* ***at***
WH THING

b. *wh-* ***ere***
WH PLACE

c. *wh-* ***en***
WH TIME
[2cm]

(176) Portmanteau interrogative pronouns in English

a. *who* (/hu:/)
WH.PERSON

b. *how* (/haʊ/)
WH.MANNER

The idea that that interrogative pronouns contain functional nominals inside is supported in a more direct way by Persian *jâ* 'place' which is morphologically contained inside *ko-jâ* 'where'. Another example which constitutes evidence for the syntactic presence of functional nouns in interrogative pronouns is an example of syncretism in Lakhota between the functional *táku* 'thing' and the pronoun *táku* 'what'.

The last category of bases consists of generic pronouns derived from the numeral *one*. The use and semantics of these forms suggests that they contain functional nominals representing the categories of THING and PERSON. The following sentences show the use of the *one* pronoun in English:

(177) a. *The drawing I made of Hollywood when I was a kid, the* ***one*** (THING) *that inspired my dad.*

b. *We need a place to sleep tonight. Let's buy that* ***one*** (THING).

c. *He is the* ***one*** (PERSON) *who embarrassed himself.*

d. *No child ever loved you, so why should this* ***one*** (PERSON) *?!*

We can also see the use of *one* as a pronoun in French. The form *l'un* 'the one' is used in a way that is similar to English *one*. *Un* is used as a base in *quelqu'-un* 'someone':[56]

(178) a. *Vous devrez passer* ***l'un*** *de nos examens.*
you.PL have.to pass one of our exams
'You'll have to take **one** (THING) of our tests.'

b. *J'espère assister à* ***l'une*** *de vos conférences.*
I-hope attend to one of your conferences
'I hope to attend **one** (THING) of your conferences.

[56]See also Jenkins (1961: 126).

c. ***L'une*** *d'entre elles a manqué le bus.*
one of them have missed the bus
'One (PERSON) of them missed the bus.'

Italian has a *one*-pronoun as well. The pronoun *uno* can be used to represent the categories of PERSON or THING:

(179) Italian

a. *Ne ho comprato* ***uno*** *stamattina.*
of.it have bought one this.morning
'I bought **one** (THING) this morning.'

b. *Ho incontrato* ***uno*** *che ti conosce.*
have met one that you knows
'I met **one** (PERSON) who knows you.'

Given the fact that generic *one*-pronouns are syncretic with the numeral *one*, it can be argued that they lexicalize functional nouns together with *one* as a nominal modifier. Therefore *one* as a pronoun will correspond to *one* THING or *one* PERSON.

The examples discussed in this section allow us to argue that bases used to form assertive indefinite pronoun syntactically contain functional nouns representing ontological categories. It is the presence of these functional nouns that determines the category of the referent that is modified by the indefinite marker. Due to their phonological realizations, nominal bases can be considered to be structurally close to functional nominals, whereas interrogative pronoun bases will realize functional nouns either together with the interrogative structure (portmanteau) or as a separate morpheme merged with an interrogative morpheme. Bases of the third category, that is generic *one*-pronouns, are syncretic with the numeral *one*, which is why they can be analyzed as functional nominals lexicalized together with numeral modifiers.

5.4. Merging indefinite markers and their bases

This section addresses aspects of merging the indefinite hierarchy with the three established types of bases. I will discuss the syntactic positioning of the indefinite hierarchy with respect to the base and the formation of indefinite prefixes.

As argued in Sections 5.2 and 5.3, the indefinite hierarchy can be projected in combination with three structurally different types of bases, that is generic nouns, interrogative/relative pronouns and generic *one*-pronouns. This can be represented the following way:

(180) a. [Indef [$NP_{functional}$]]
b. [Indef [Wh [$NP_{functional}$]]]
c. [Indef [$Pron_{one}$]]

This representation, in which the indefinite hierarchy (Indef) is projected on top the base is suggested by instances of syncretism between interrogative pronouns and indefinite pronouns (Haspelmath 1997: 170, Dyirbal - Dixon 1972: 265, Duthc - Hengeveld et al. 2020, Khmer - Huffman 1967: 153-6, Hopi - Malotki 1979: 110):

(181) Wh-pronouns syncretic with indefinite pronoun
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*)

These observations suggest that the indefinite hierarchy can be realized together with an interrogative pronoun base as a portmanteau morpheme, which results in syncretism. The desired form of syncretism can be obtained if a wh-pronoun base and the indefinite hierarchy are lexicalized as shown in (182) on the basis of Chinese:

(182) a. [Wh-$pronoun_{THING}$] ⇒ *shénme* 'what'
b. [Indef [Wh-$pronoun_{THING}$]] ⇒ *shénme* 'something'

Another evidence for this kind of representation comes from Lakhota, where the word *táku* can be used as a functional noun 'thing', an interrogative pronoun 'what' and an indefinite pronoun 'something'. Note how this data are in line with the syntactic containment of functional nominals inside interrogative pronouns argued for in Section 5.3. Baunaz and Lander (2018a:

3) present the following representation of the Lakhota facts:

(183) Functional noun, wh-pronoun, indefinite pronoun: syncretism

a. [THING] ⇒ *táku*

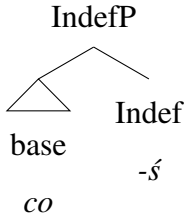
b. [Wh [THING]] ⇒ *táku*

c. [Indef [Wh [THING]]] ⇒ *táku*

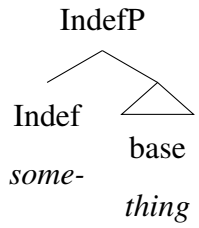
The presented examples of syncretism indicate that the indefinite hierarchy can be projected on top of a nominal base (functional noun) or on top of an interrogative base containing a functional nominal. It is likely that the indefinite hierarchy is merged with *one*-pronouns in an analogous way.

Another issue that has to be addressed before the structure and derivation of assertive indefinite markers may be properly represented is connected with the formation of prefixal indefinite markers. First, note that indefinite markers can become either suffixes (as in Polish *co-**ś*** 'something') or prefixes (as in English ***some**-thing*). This can be represented the following way:

(184) a. Suffix



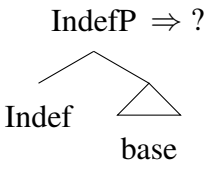
b. Prefix



The formation of an indefinite suffix is straightforward and stems from the spellout algorithm,

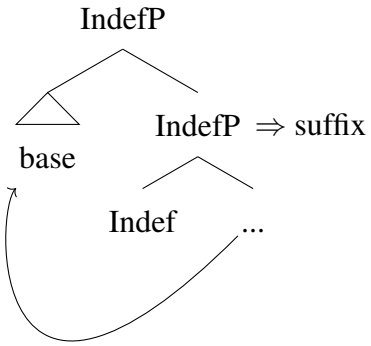
which will trigger spellout-driven movement if there is no lexical entry that can be used to lexicalize an indefinite feature (Indef) projected on top of a base:[57]

(185)



The lack of a matching lexical entry will result in the displacement of the base to a position above the indefinite feature, followed by another spellout attempt. If a matching lexical entry is found, IndefP will be lexicalized as a suffix:

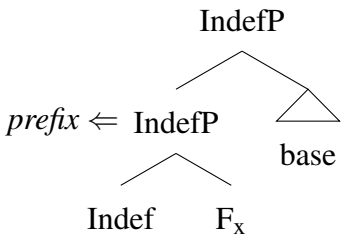
(186) Indefinite suffix



In contrast, the formation of an indefinite prefix will begin as the last stem in the spellout algorithm (see Chapter (116)). This will occur if a matching lexical entry cannot be found after spellout-driven movement and backtracking followed by spellout-driven movement. When all other options fail, a subderivation will be spawned in order to lexicalize the newly merged feature (Indef in this case). As mentioned in Section 4.3, a subderivation will begin with the merger of two features, which will result in a binary structure at the bottom of the subderivation. The first feature provided in the subderivation (here labeled F_x) will serve as the base for the merger of Indef:

[57] Here, I use the label Indef to represent the first feature of the indefinite hierarchy (F_1).

(187) Prefix formation

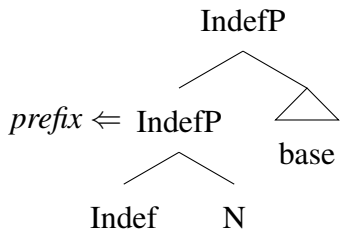


There is no clear consensus regarding the base feature that has to be provided for the initial merger in a subderivation. A possible solution is proposed in Caha et al. (2019), where the first feature used to form the binary bottom of a subderived structure is a copy of the last successfully merged feature from the main spine. I will not follow this particular proposal due to the fact that the indefinite hierarchy can be projected on top of three structurally different bases. We can see this in English:

(188) a. *some-one* (one-pronoun base)
b. *some-thing* (nominal base)
c. *some-where* (pronominal base)

If the prefix (*some-*) remains structurally uniform in combination with different types of bases, which is indicated by the fact that its phonological exponent does not change, the proposal discussed in Caha et al. (2019) would require any suitable base to end with the same feature merged at the top node. Only then would it be possible to begin a subderivation with the same feature as the base for the merger of the indefinite hierarchy. Since it seems unlikely that the three base types have the same feature as the last element in their structure, I will instead make a proposal based on suggestions made in Starke (2018) and Wiland (2019). To form the bottom binary structure of an indefinite prefix, the derivation system will first provide a nominal base feature (N) and merge it with Indef. The nominal feature can become a base for the merger of the indefinite hierarchy since it reflects the nominal nature of the three types of bases that are used to form indefinite pronouns:

(189) Indefinite prefix



Once the provided indefinite feature has been successfully lexicalized, the subderivation will be closed and merged with the base, which will result in the formaition of a prefixal indefinite marker.

In the following sections, I will demonstrate on the basis of selectged languages how the attested forms of syncretism can be obtained when the indefinite hierarchy is matched by different lexical entries. The proposed representations will also illustrate the derivation of indefinite markers as suffixes and prefixes in accordance with the descriptions provided in this section (see (186) and (189)).

5.4.1. Russian: no syncretism

Before any of the patterns of syncretism can be tackled, we should take a look at languages in which we see no syncretism. In languages of this kind, each subset of the indefinite hierarchy is spelled out with a separate lexical entry, which results in three different markers expressing the three indefinite functions. One of such languages is Russian, in which the three markers are *-nibud* (non-specific), *-to* (specific unknown) and *koe-* (specific known). This means that the paradigm for Russian is as follows:

Table 10. Russian: no syncretism.

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

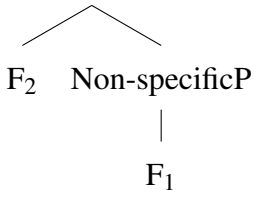
The paradigm shown in Table 10 is the consequence of the indefinite hierarchy being spelled out by three separate lexical entries. Each entry is a match for only a single subset of the hierarchy. The three lexical entries used to lexicalize the indefinite hierarchy in Russian are

given below:

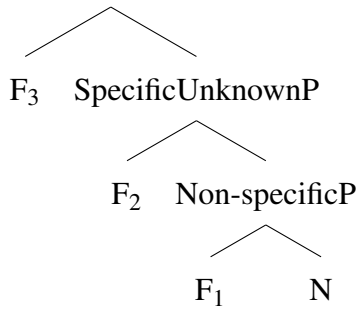
(190) Lexical entries

a. Non-specificP $\Leftrightarrow$ *-nibud*

b. SpecificUnknownP $\Leftrightarrow$ *-to*



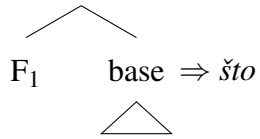
c. SpecificKnownP $\Leftrightarrow$ *koe-*



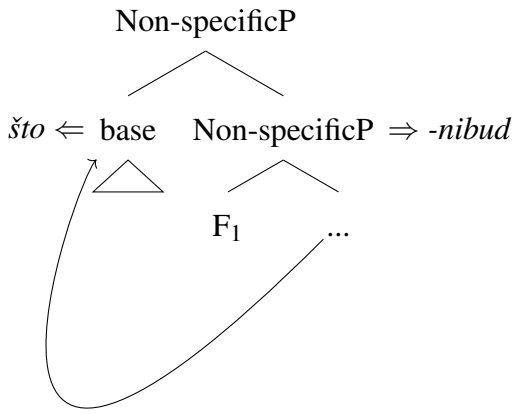
Two of the three indefinite markers in Russian (the non-specific and specific unknown markers) are suffixes, which means that they are derived through spellout-driven movement involving displacement of the base. For all trees in this section, I will use the base corresponding to the category of THING, namely *što* 'what'.

When F_1 is merged with the base and cannot be spelled out with any of the available lexical entries, the spellout system will initiate movement (roll-up) of the base to the position above F_1. The displaced base will not leave a trace and will not project a new phrase at the top of the tree. The movement will create a constituent matching the lexical entry in . The constituent F_1P will be lecicalizaed as *-nibud*:

(191) Non-specificP ⇒ *?*



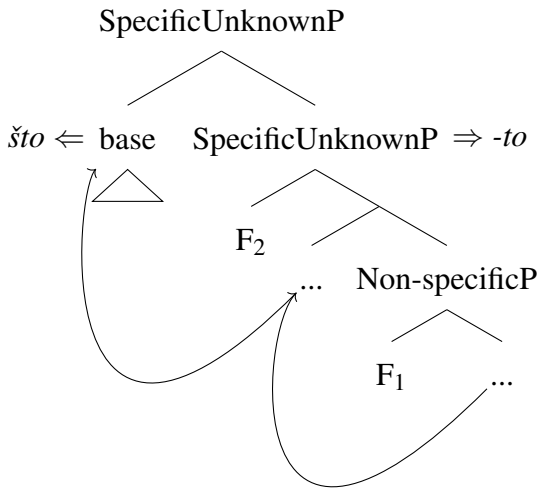
(192) Non-specific suffix



The merger of the next feature in the hierarchy (F_2) will not result in successful spellout either. Following the spellout algorithm, the lexicalization system will again trigger movement of the base to obtain a structure that can be matched with one of the available lexical entries.[58] As in the derivation of the non-specific marker, the displaced base will not leave a trace or project. A constituent containing F_1 and F_2 can be spelled out with entry as *-to*:

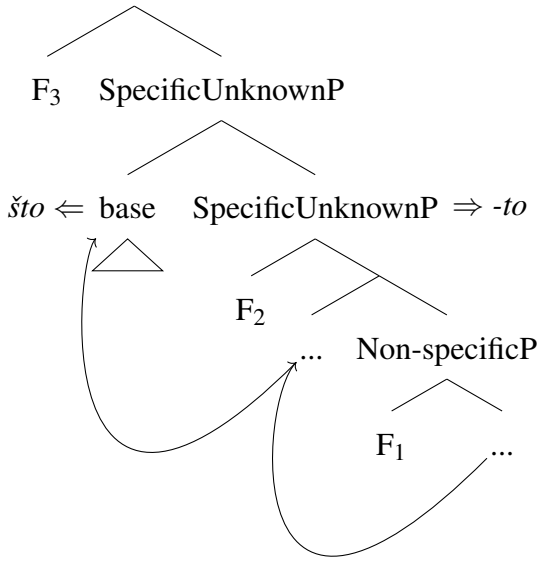
[58]The first step of the spellout algorithm.

(193) Specific unknown suffix



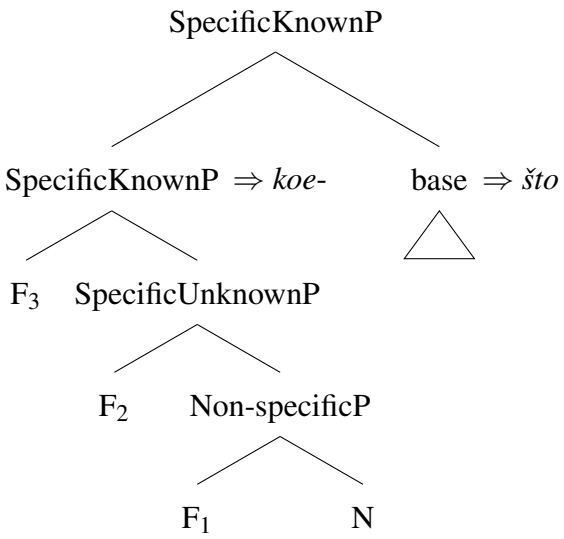
The prefixal marker *koe-* requires a more complex derivation than the other two. The last feature in the sequence (F_3) cannot be spelled out right after merge or after movement of the base. Even backtracking will not produce a tree geometry that can be matched with the lexical entry containing F_3 :

(194) SpecificKnownP ⇒ *?*



Since the specific-known marker equals the whole indefinite hierarchy, the Russian *koe-* has to be inserted as the phonological exponent of a structure containing the sequence of F_1, F_2, F_3. Additionally, *koe-* is a prefix, which means that it is derived through subderivation. To form a left branch constituent containing all three indefinite features, the system has to backtrack all the way to the base and trigger subderivation at that point. A new derivation space will be formed in order to create a lexicalizable XP which can provide F_3. Of course, to merge F_3, other features in the hierarchy, that is F_1 and F_2 will have to be derived first. Additionally, a base feature N will appear at the bottom of the subderivation to form the minimal binary structure together with F_1.[59] The subderivation will be kept active until the whole indefinite hierarchy has been merged and lexicalized with the lexical entry in . Once there are no more features to merge and spell out, the subderived phrase will be integrated into the main derivation as a left branch (prefix):

(195) Specific known prefix



5.4.2. Lithuanian: no syncretism

Lithuanian is another language in which no syncretism between the three types of indefinite markers is observed. The three indefinite markers used in Lithuanian are *-nors* (non-specific), *kaž-* (specific unknown) and *kai-* (specific known):

[59] See Section 5.4.

Table 11. Lithuanian: no syncretism.

	non-specific	specific unknown	specific known	pattern
Lithuanian	-nors	kaž-	kai-	ABC

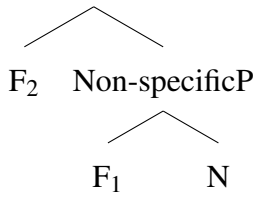
As in the case of Russian, the lack of syncretism is the consequence of the indefinite hierarchy being spelled out with three separate lexical entries. These three entries are as follows:

(196) Lexical entries

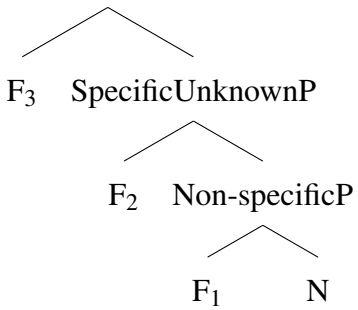
a. Non-specificP $\Leftrightarrow$ *-nors*

F_1

b. SpecificUnknownP $\Leftrightarrow$ *kaž-*

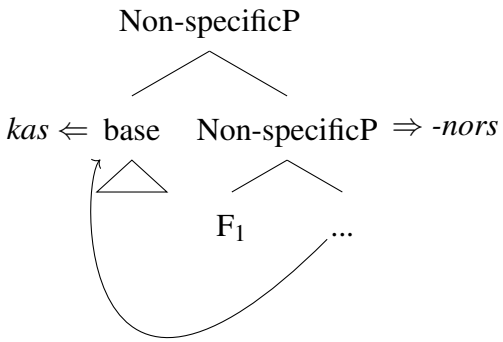


c. SpecificKnownP $\Leftrightarrow$ *kai-*



Only the non-specific marker is a suffix, which means that only this indefinite marker is derived through roll-up movement of the the structure below F_1. The displacement of the base (*kas* 'what') will result in a tree geometry in which F_1P constitutes a constituent for which (196-a) is a matching lexical entry:

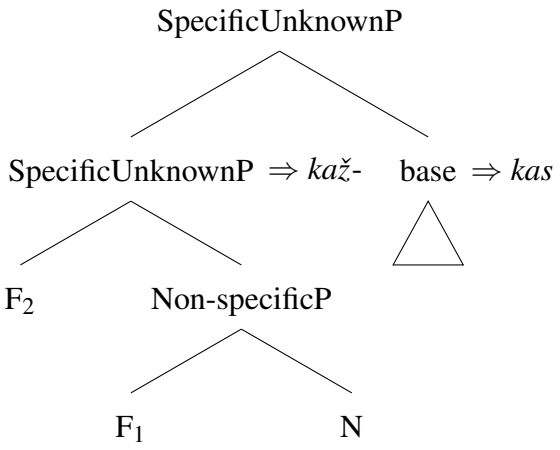
(197) Non-specific suffix



The other two indefinite markers, that is the specific unknown and specific known markers, are prefixes. As mentioned in Section 4.3, the derivation of a prefix involves a subderivation creating a left-branch constituent. Hence, when F_2 is merged with the structure shown in (197), no matching lexical entry will be found in the lexicon, which means that the lexicalization system will attempt to alter the derived structure in order to obtain a lexicalizable configuration. As with the derivation of the Russian specific known marker, the system will go through the steps of the spellout algorithm, until it can backtrack to the base and spawn a subderivation. For the formation of the specific unknown marker, the subderivation will first merge F_1 and a base nominal feature N as the initial binary structure. This structure will be lexicalized as a subset of the entry in (196-b). Once F_2 is added to the subderivation in the next merge-spellout cycle, the structure will become a direct match for the entry in (196-b):[60]

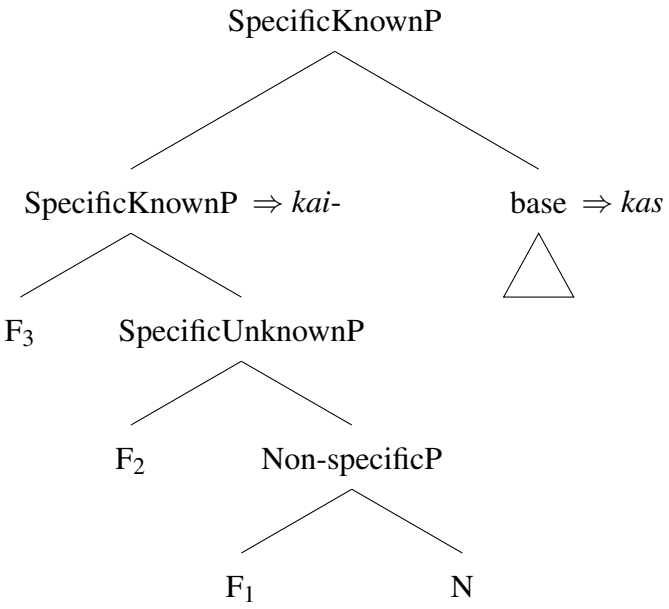
[60]The entry in (196-c) will not become a match due to the Elsewhere Principle.

(198) Specific unknown prefix



As for the derivation of the specific known marker, F_3 can be provided and lexicalized inside the subderivation spawned during the derivation of the specific unknown marker. The subderivation will be kept active and simply provide the next feature in the sequence (F_3). The resulting structure will match the lexical entry in (196-c), which means that it will be spelled out as *kai-*. Since there are no further features to merge in the indefinite hierarchy, the subderivation will be closed and merged with the main structure:

(199) Specific known prefix



5.4.3. English: AAA pattern

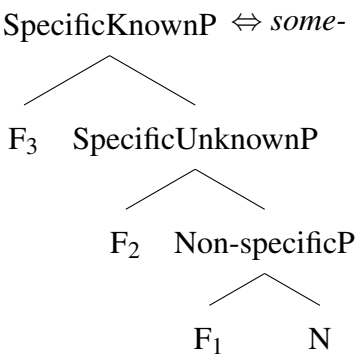
English is one of the languages that exemplifies the AAA pattern, which means that the non-specific, specific unknown and specific known are syncretic. In other words, the syntactic structures corresponding to the non-specific, specific unknown and specific known indefinite functions are realized by the same phonological exponent *some-*. Hence, the paradigm for English is as follows:

Table 12. English: AAA pattern.

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

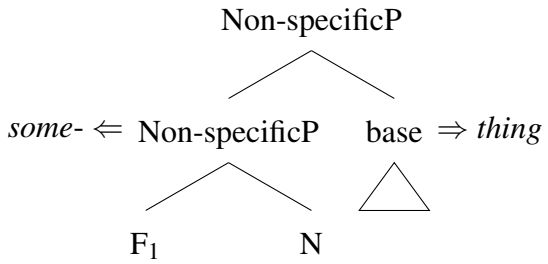
The three markers in English are also all prefixes, which means that they constitute subderived phrases merged with the base as complex left branches. The initial attempt (after f-merge) to spell out the first layer of the indefinite hierarchy (F_1) will fail, and will be followed by spellout-driven movement. Since spellout-driven movement will also not generate a lexicalizable tree geometry (even after backtracking), the final last-resort operation will trigger and derive a prefix. As for the derivation of the two next layers of the hierarchy (F_2 and F_3), these features can also be spelled out inside the subderivation. All three layers of the indefinite hierarchy will be spelled out with only a single lexical entry, which will result in the observed syncretism (AAA). Example (200) shows the necessary lexical entry:

(200) Lexical entry



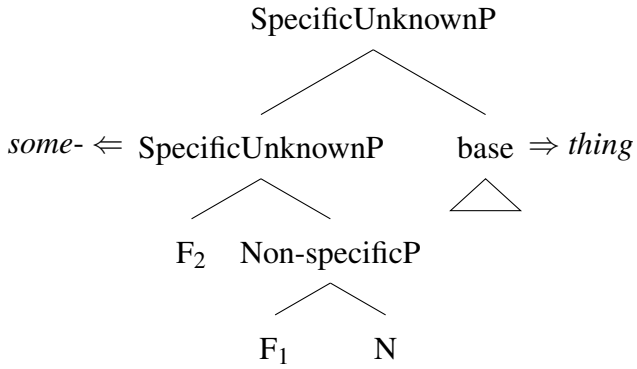
As already mentioned, the need to spell out F_1 will lead to the formation of a subderivation. The new phrase will be generated through the merger of F_1 with a nominal base feature (N). This will result in the creation of a structure that is a subset of the lexical entry in (200), which means that it can be successfully spelled out and integrated into the main derivation:

(201) Non-specific prefix



The remaining layers of the indefinite hierarchy, that is F_2 and F_3 will also be spelled out as parts of the subderived constituent. Instead of closing the subderivation and integrating it into the main structure immediately after F_1 has been provided, the system may merge (and lexicalize) more features. If F_2 and F_3 are available for merge, these features will be added to the subderived structure. The insertion of F_2 will again result in a constituent for which (200) is a matching lexical entry (under the Superset Principle). The outcome of the first lexicalization cycle will be overwritten by the same lexical exponent (*some-*):

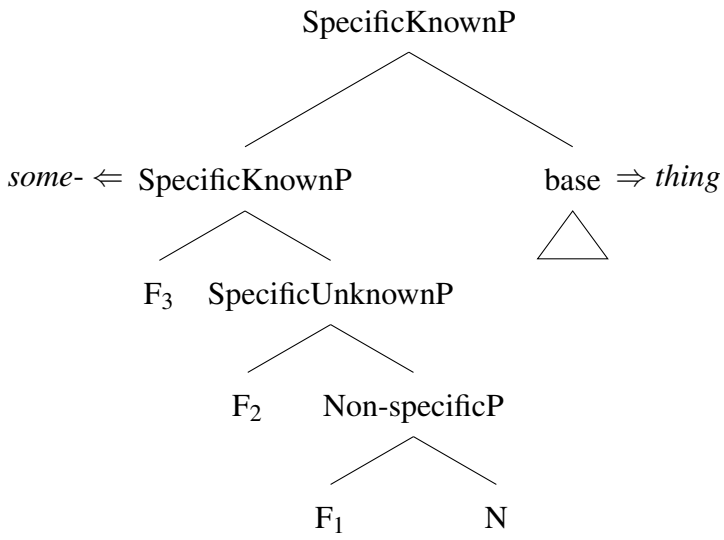
(202) Specific unknown prefix



The same outcome will be obtained when F_3 is merged with the subderivation. The structure will once again be spelled out as *some-*. The spellout of all three layers of the indefinite hierarchy with a single lexical entry will result in total syncretism of indefinite markers (the

AAA pattern):

(203) Specific known prefix



5.4.4. Polish: AAA pattern

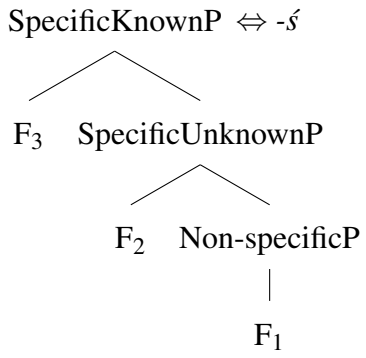
Polish exemplifies another variant of the AAA pattern due to the fact that all indefinite markers (non-specific, specific unknown and specific known) are realized as suffixes (*-ś*). This means that the three indefinite markers in Polish are derived through multiple movements of the base. The indefinite marker paradigm for Polish is as follows:

Table 13. Polish: AAA pattern.

	non-specific	specific unknown	specific known	pattern
Polish	-ś	-ś	-ś	AAA

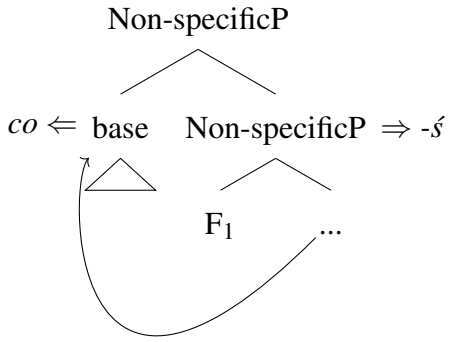
As in the case of English, the indefinite hierarchy in Polish is spelled out with only a single lexical entry. This lexical entry is shown in example (204) below:

(204) Lexical entry



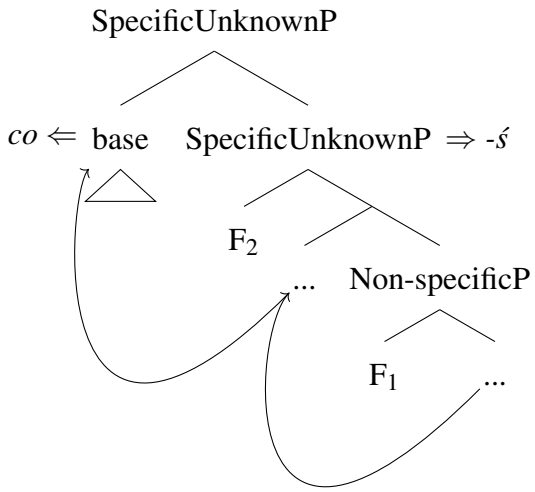
The merger of F_1 with the base will not result in immediate spellout since there is no lexical entry in the lexicon that ca become a match for such a structure. Subsequently, the spellout system will attempt to displace a piece of the structure to obtain a lexicalizable tree geometry. In this case, a lexicalizable structure can be obtained through the displacement of the complement of F_1 (roll-up of the base). The stranded feature may then be spelled out since the entry in (204) becomes a match under the Superset Principle:

(205) Non-specific suffix



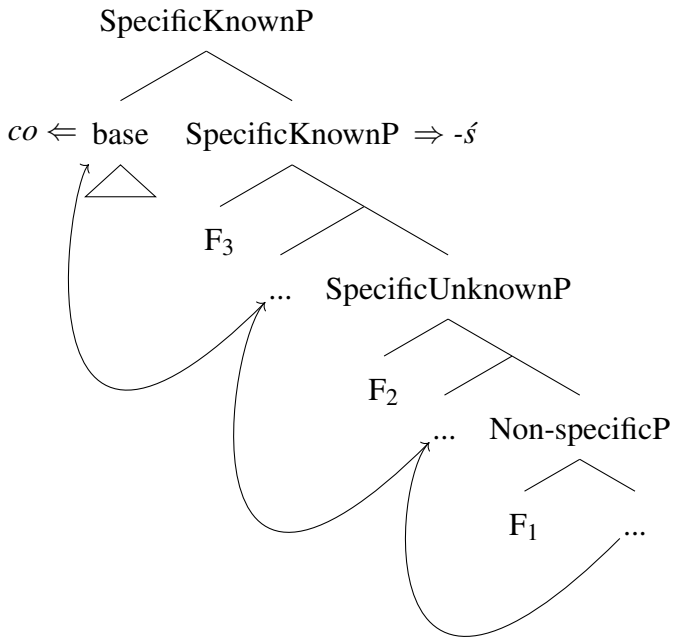
After F_2 has been added to the structure, the scenario described above will repeat. The base will be displaced again, which will form a lexicalizable constituent containing F_1 and F_2:

(206) Specific unknown suffix



The same steps will be repeated for F_3. The base will be displaced again, which will lead to the formation of a constituent consisting of F_1, F_2 and F_3. This constituent will again be lexicalized with the lexical entry in (204). Thus, the whole indefinite hierarchy, as well as its subsets, are lexicalized with a single lexical entry, which leads to the emergence of the AAA pattern:

(207) Specific known suffix



5.4.5. Georgian, Ossetic and Yakut: ABB pattern

The ABB pettern is exemplified by three languages from the studied languages sample, namely Georgian, Ossetic and Yakut. In these languages, the specific indefinite markers are syncretic to the exclusion of the non-specific marker:

Table 14. Georgian, Ossetic, Yakut: ABB pattern.

	non-specific	specific unknown	specific known	pattern
Georgian	-me	-γac	-γac	ABB
Ossetic	-ty	-dær	-dær	ABB
Yakut	-eme	-ere	-ere	ABB

As it can be seen above, the three indefinite markers are all suffixes in all three languages. This results from the displacement of the stem following the lack of successful spellout after f-merge. In all three cases, the emergent ABB pattern is the consequence of two lexical entries spelling out the indefinite hierarchy. One lexical entry is a match for the non-specific structure,

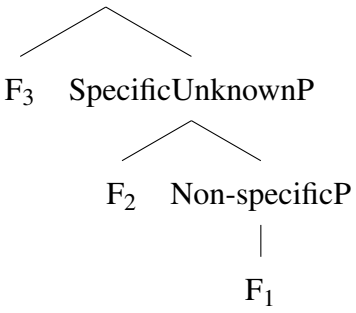
and the other is used to lexicalize the specific unknown and specific known structures. I will use Georgian to illustrate the derivation of the ABB pattern. The lexical entries used to lexicalize the indefinite hierarchy in Georgian are shown in the example below:

(208) Lexical entries

a. Non-specificP $\Leftrightarrow$ *-me*

F_1

b. SpecificKnownP $\Leftrightarrow$ *-γac*



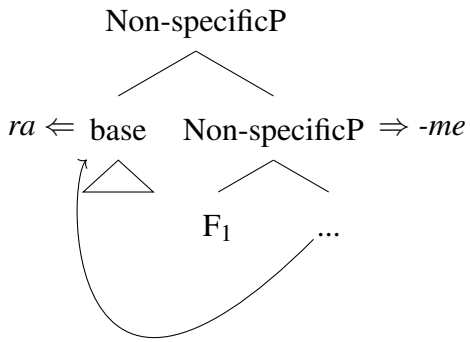
F_3 SpecificUnknownP

F_2 Non-specificP

F_1

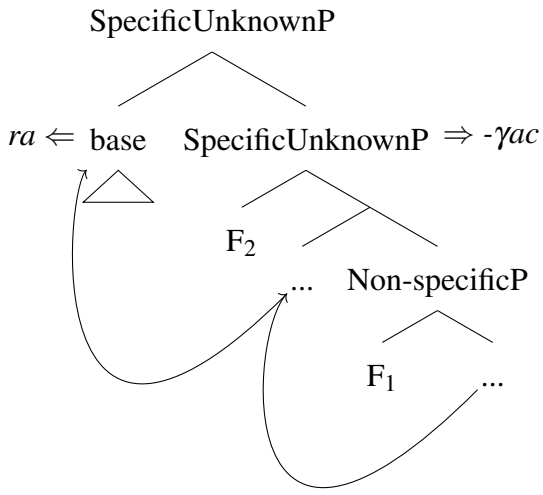
The non-specific indefinite marker is derived when the base is rolled up, which is the first last-resort option available after the initial attempt (right after f-merge) to spell out F_1 together with the stem has failed. Even though both lexical entries in (208) are potential matches for the resulting structure, the lexical entry in will be chosen over due to the fact that it does not contain any superfluous features (the Elsewhere Principle):

(209) Non-specific suffix

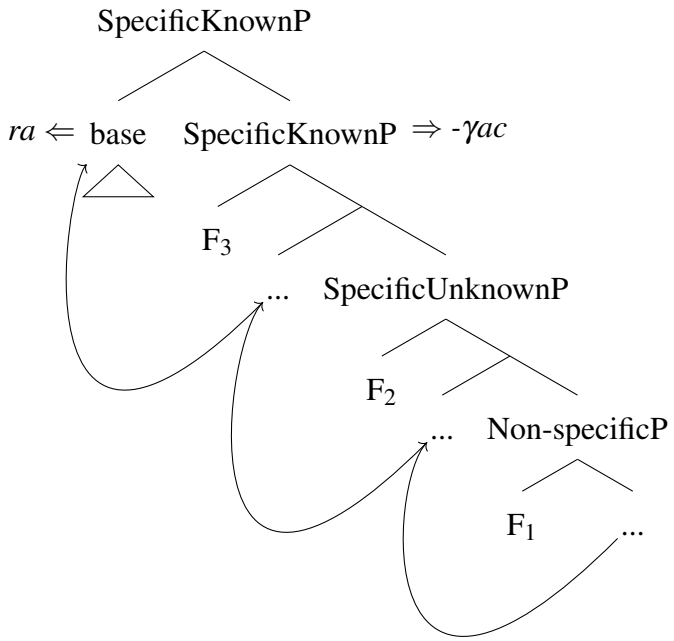


The lexical entry in (208) will become a match for the specific unknown and specific known indefinite structures as permitted by the Superset Principle. In both cases, the lexicalizable tree geometry is also achieved through the displacement of the base to a higher position. The ABB pattern is the consequence of the non-specific structure receiving a different phonological exponent than the other two subset of the indefinite hierarchy:

(210) Specific unknown suffix



(211) Specific known suffix



5.4.6. Latin: AAB pattern

The AAB pattern is found in Latin, where the non-specific and specific unknown markers are syncretic to the exclusion of the specific known marker. The two phonological exponents used in Latin are *ali-* (non-specific and specific unknown) and *-dam* (specific known):

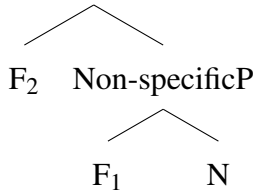
Table 15. Latin: AAB pattern.

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

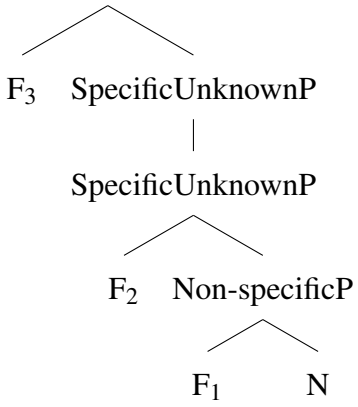
As in the case of languages where the ABB pattern is observed, the emergence of the AAB pattern is connected with the presence of two lexical entries that can be matched with the indefinite hierarchy. The lexical entries available in Latin are shown below:

(212) Lexical entries

a. SpecificUnknownP ⇔ *ali-*

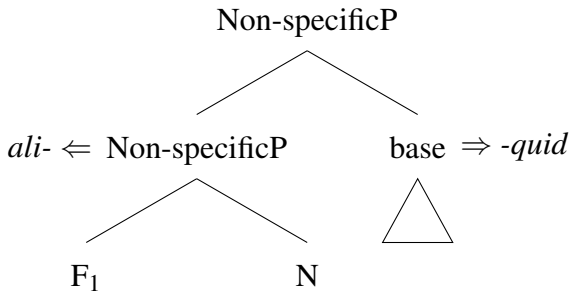


b. SpecificKnownP ⇔ *-dam*



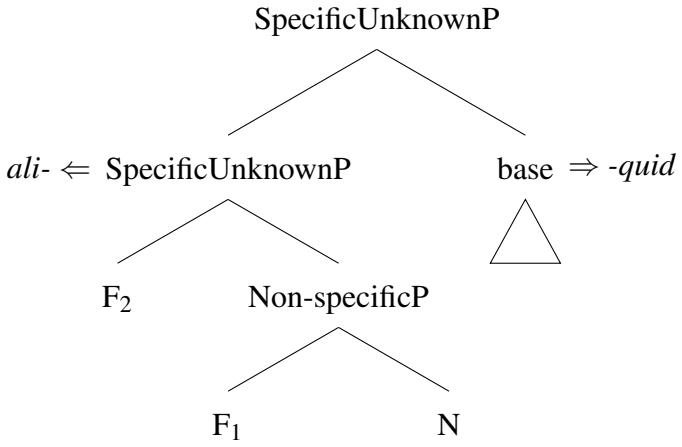
The first two indefinite markers are prefixes with respect to the base, which is the result of them being derived through the last-resort operation known as subderivation. The merger of F_1 with the base will not lead to successful spellout (due to the lack of a matching lexical entry), and the lexicalization of this feature will not be possible until the last step of the spellout algorithm (subderivation) is triggered. The subderivation will form the minimal binary structure by merging F_1 with a nominal feature (N). Subsequently, the resulting phrase will be spelled out as a subset of the lexical entry in (212-a) and integrated into the main spine:

(213) Non-specific prefix



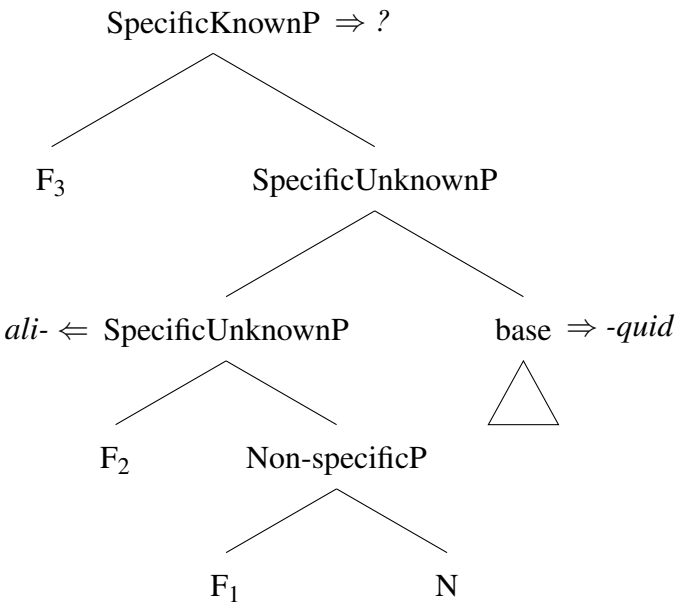
To form the specific unknown structure (F_1, F_2), the syntactic system has to add F_2 to the subderivation through regular f-merge. Merging F_2 on top of the main derivation will not produce a lexicalizable structure (even after transformations), which is why the subderivation has to remain active and provide the desired feature. Once F_2 has been added, the subderivation, now containing F_0, F_1 and F_2, will be spelled out with the lexical entry in (212-a) as *ali-*:

(214) Specific unknown prefix



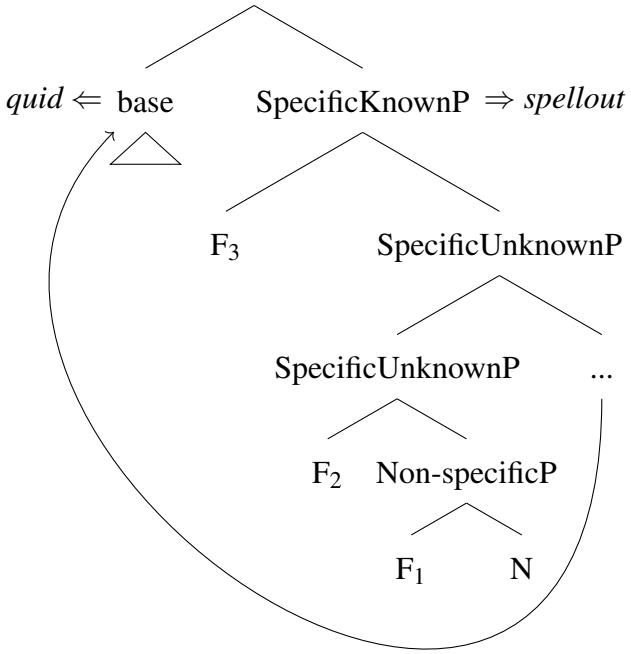
The specific known structure in Latin corresponds to the morpheme *-dam*. In contrast with the other two indefinite markers, the specific known marker in Latin is a suffix. This fact should make it quite clear at this point that merging F_3 on top of the derivation in (214) will not provide a lexicalizable tree geometry:

(215)



The derivation of the Latin specific known marker, which is an example of a prefix (*ali-*) turning into a suffix (*-dam*), poses a problem for the current spellout algorithm. This is because none of the steps in the spellout algorithm (as described so far) applied to the structure in (215) will create a tree geometry where F_1, F_2 and F_3 form a constituent that can be spelled out as a suffix with respect to the base. The simplest way to obtain the desired structure would be to displace the base constituent to a position above F_3, which is however not permissible under the spellout algorithm:

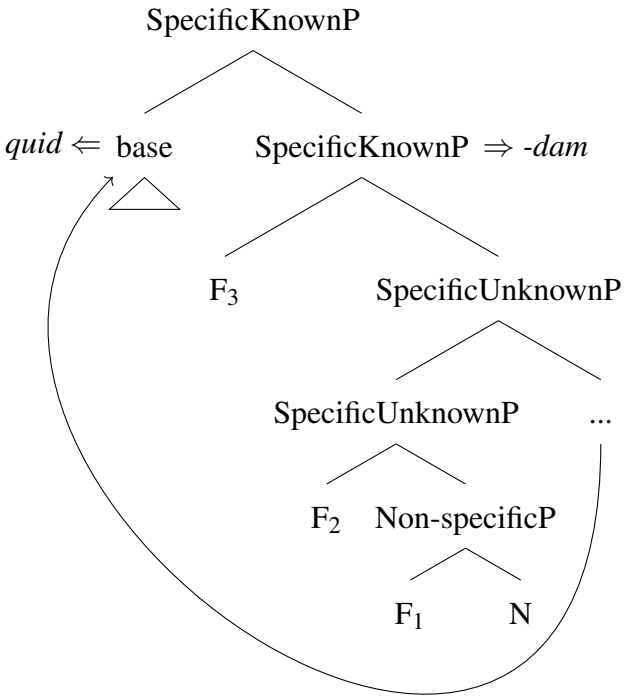
(216) Stem extraction



Confronted with this obstacle, I suggest that we should consider the possibility that the spellout algorithm consists of more steps than cyclic movement, roll-up, backtracking and subderivation. The case of Latin suggests that there may be another option between the first two steps of the algorithm (spec-to-spec and roll-up movements) and backtracking, namely subextraction. It would target for movement the complement of the previously inserted node

and displace it to the position above the the feature that has been merged in the current derivation cycle. This kind of spellout-driven movement would allow us to provide a simple solution to the problem of *-dam* being a suffix. Once the base (*quid*) is extracted, the remaining constituent will be spelled out with the lexical entry in (212-b):[61]

(217) Spellout of F_3P (*-dam*)



The AAB pattern appears to be quite uncommon among other patterns of indefinite marker syncretism. The studied languages sample contains only one language in which this pattern can be seen, namely Latin. This may be connected with the fact that the derivation of the AAB pattern requires subextraction from a previously formed constituent (the complement of the newly provided feature), which may be considered a more complex operation than other forms of spellout-driven movement (spec-to-spec and snowball). Additionally, constituents

[61] Note that the displacement of the base to the top of the structure will not remove the node projected by the previously derived prefix (F_2P).

formed through subextraction, with a gap created by the evacuation of a piece of structure, will require complex lexical entries to be successfully lexicalized, which may also contribute to the rarity of the AAB pattern. As remarked in Bobaljik (2012), the AAB pattern is rather rare and difficult to derive with the use of the mathodological framework employed by the author (Distributed Morphology). This supports the predictions that stem form the studied indefinite pronoun data sample and the subsequent analysis thereof.

5.5. Portmanteau and wh-pronoun indefinites

In a number of languages, we can observe examples of indefinites that take the form of bare interrogative pronouns. Such indefinites often realize to one of the three discussed assertive indefinite functions (non-specific, specific unknown or specific known). Consider some examples:

(218) Chinese *shei* 'who' (Zhuo Chen: 2017)

Ruguo ***shei*** *qifu ni, jiu gaosu wo*.
if who bully you then tell me
'If **somebody** bullies you, tell me.'

(219) German *wo* 'where' (Postma 1994: 192)

Er hat ***wo*** *gewohnt*.
he has where lived
'He has lived **somewhere**.'

(220) (Dutch *wat* 'what' (Postma 1994: 193)

Jan heeft snel ***wat*** *opgeschreven*.
John has quickly what written
'Jan has quickly written **something**.'

(221) Passamaquoddy (Bruening 2007:153)

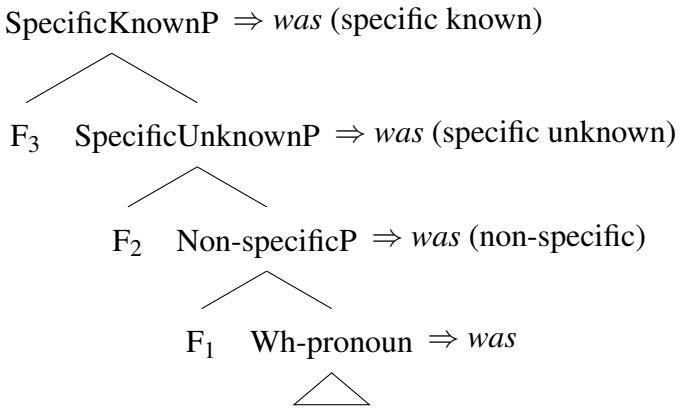
Itom ***wen-il*** *nemiy-a-t*?

say.3 who-OBV IC.see-DIR-3CONJ5

'Did he say he saw **someone**?'

The indefinite pronouns used in the examples above are phonologically identical to wh-pronouns, which is why it is not possible to identify the indefinite marker and the pronoun base. The nanosyntactic framework can however easily explain the existence of wh-indefinites and their syntactic structure. Indefinite pronouns that share their form with interrogative pronouns constitute examples of syncretism. What this means is that interrogative pronouns and their corresponding indefinites may be spelled out with the same lexical entry. This is made possible by the fact that wh-pronouns, as one of the available indefinite pronoun base types, may constitute syntactic subsets of indefinite pronouns (see Section 5.4). In other words, interrogative pronouns and indefinites that have wh-pronoun as their bases form a containment relation. Consider, the following lexicalization of German *was*, which may be either a wh-pronoun 'what' or an indefinite pronoun used in the non-specific, specific unknown or specific known functions 'something':

(222) German: *was* 'what'/'something'

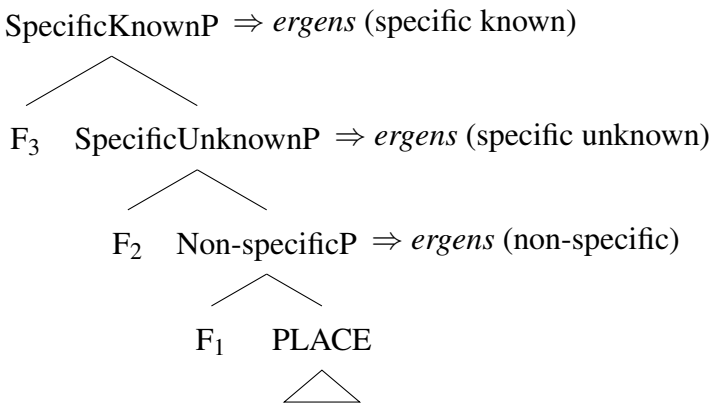


First, the wh-pronoun is derived and lexicalized under the Superset Principle as a subconstituent within a lexically stored tree matching the whole structure in (222). Subsequently, the indefinite hierarchy is built with the wh-pronoun as the base. With each layer of the indefinite structure, a larger subset of the lexical entry becomes a match. Thus, the interrogative pronoun and the three indefinite pronouns (non-specific, specific unknown and specific known) are spelled out

with a single lexical entry, which leads to the observed syncretism.

A similar analysis may be applied to other indefinites that do not follow the marker + base pattern, such as monomorphemic indefinites. Examples of this kind can be found in Dutch, where the pronouns *ergens* (somewhere) and *ooit* (sometime) no longer seem to be divisible into parts.[62] In cases like these, the indefinite hierarchy is also spelled out together with the base. This is shown in the following representation:

(223) Dutch: *ergens* 'somewhere'



The base expressing the category of PLACE will be embedded in the structure of the indefinite and will not surface as a separate morpheme. In the course of the derivation, the base will be spelled out either with the same lexical entry as the indefinite structure or another lexical entry that will become a better match. In the latter case, the phonological exponent associated with that entry will be overwritten by *ergens* when the indefinite hierarchy is built on top of the base.

5.6. Cumulative vs. gapped structure

In the proposed analysis of non-specific, specific unknown and specific known indefinite markers, I make a case for a syntactic hierarchy of indefinite features. According to this model, each of the three types of indefinite markers corresponds to a different constituent based on the proposed indefinite hierarchy. Therefore, the non-specific, specific unknown and specific known indefinite functions are expressed by the following sets of features:

[62] *Ergens* appears to originate from Proto-Germanic *ajw 'ever' + *hwar-gin (where-PT), while ooit may come from *ajw + jet 'still' (Haspelmath 1997: 246).

(224) a. $[F_1] \Rightarrow$ Non-specific function

b. $[F_2\ [F_1]] \Rightarrow$ Specific unknown function

c. $[F_3\ [F_2\ [F_1]]] \Rightarrow$ Specific known function

One of the most crucial aspects of the way the indefinite categories are represented is that the non-specific, specific unknown and specific known indefinite markers are formed through cumulation (stacking) of syntactic features. Three non-identical sets of features, and consequently, three separate syntactic structures, are obtained when the derivation is grown by merging new features. Since the sequence according to which the indefinite hierarchy is assembled is taken to be cross-linguistically constant, the indefinite features forming the hierarchy follow a strict and fixed relative order. In consequence, a layer of the hierarchy is available for f-merge only when the preceding item (specified by the functional sequence) is present in the structure. In other words, the proposed model predicts that if F_n has not been merged, F_{n+1} cannot be provided for f-merge.

However, another representation the non-specific, specific unknown and specific known indefinite markers may seem possible, namely a non-cumulative model. In this kind of representation, the syntactic structures corresponding to the non-specific, specific unknown and specific known indefinite markers do not form a hierarchy based on syntactic containment and feature cumulation. Three non-identical sets of features can be obtained in this way if indefinite features do not have to be assembled in a strict order and the features corresponding to one indefinite marker type may be absent from the structure of another. Below, I evaluate this alternative model in the context of the collected data concerning syncretism. Ultimately, the non-cumulative representation has to be rejected in favor of the cumulative model due to the fact that it does not allow us to obtain the attested patterns of syncretism under one of the core assumptions of the nanosyntactic system of lexicalization, namely the Superset Principle (as formulated in Section 4.2).

First, consider again how cyclic lexicalization (cyclic override) in the cumulative model leads to indefinite marker syncretism. This will be shown on the basis of the following data from Georgian. Table 16 shows the three sets of features (formed through feature cumulations) that give rise to the non-specific, specific unknown and specific known indefinite markers. As already discussed in Section 5.4.5, in Georgian, we observe syncretism between the specific unknown and the specific known markers (-γac) to the exclusion of the non-specific marker

(-me):

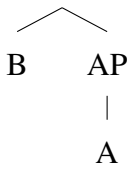
Table 16: Indefinite features: cumulation

	F_1	F_2	F_3
a. *me* (non-specific)			
b. *γac* (specific unknown)			
c. *γac* (specific known)			

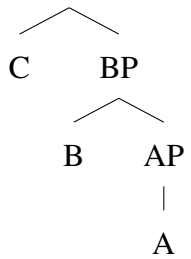
Due to the fact that in Georgian only two lexical entries are used to lexicalize the indefinite hierarchy, two sets of features, [F_1, F_2] and [F_1, F_2, F_3], will receive the same phonological exponent (-*γac*). This permitted by the Superset Principle, which allows a lexical entry to spell out trees that it is overspecified for. The smallest subconstituent of the hierarchy, that is [F_1], will receive a different exponent since there is a lexical entry that constitutes a direct match for this structure. As mentioned in Section 4.2, the Elsewhere Principle guarantees that a lexical entry with the smallest number of superfluous features will always be preferred over other matching lexical entries. In the case of Georgian, this means that the lexical entry corresponding to the exponent *-me* will always win the competition with the other lexical entry (-*γac*) (see Section 5.4.5):

(225) a. AP ⇒ me

b. BP ⇒ γac



c. CP ⇒ γac



As argued in Chapter 5, under the assumptions of the nanosyntactic system of lexicalization such as the Superset Principle and the Elsewhere Principle, a model in which features of the indefinite hierarchy are merged cumulatively allows us to properly obtain the patterns of syncretism attested in the studied language sample and account for the absence of of the ABA pattern. However, the same cannot be said about the non-cumulative representation.

A non-cumulative model of derivation for the non-specific, specific unknown and specific known markers means that three non-identical sets of features, each corresponding to a

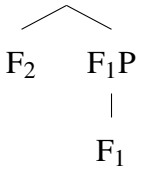
different type of marker, may be obtained in a way that does not involve feature cumulation. Consider the following scenario based on the syncretism between the specific unknown and specific known markers in Georgian. Assuming that the non-specific marker corresponds to F_1, the structural difference between the specific unknown and specific known indefinite markers may be achieved through a gap in the sequence of [F_1, F_2, F_3]. Two non-identical structures can be formed if F_2 is found in the structure of the specific unknown marker but absent from the structure of the specific known marker. In Table 17, as well as in the representations in (226), the specific unknown marker lexicalizes F_1 and F_2, while the specific known marker corresponds to F_1 and F_3. Consequently, the two indefinite markers are structurally different and not formed through feature cumulation:

Table 17. Gapped sequence 1.

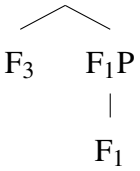
	F_1	F_2	F_3
a. *γac* (specific unknown)			
b. *γac* (specific known)			

The tree representations based on Table 17 are as follows:

(226) a. $F_2P \Rightarrow \gamma ac$



b. $F_3P \Rightarrow \gamma ac$



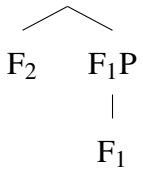
There is another scenario we may consider. This time, the structural difference between the specific unknown and the specific known markers is obtained by skipping F_1 in the structure corresponding to the specific known function. In consequence, the specific unknown marker would consist of F_1 and F_2, while the specific known marker would contain F_2 and F_3. As in the previous case, the non-specific marker corresponds to only a single feature (F_1). This is shown in the following lexicalization table and representations:

Table 18. Gapped sequence 2.

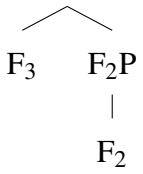
	F_1	F_2	F_3
a. *γac* (specific unknown)			
b. *γac* (specific known)			

The trees below are based on Table 18:

(227) a. $F_2P \Rightarrow \gamma ac$



b. $F_3P \Rightarrow \gamma ac$



Consider one more example where the difference between the non-specific and specific unknown indefinite markers is obtained in a non-cumulative way. In Latin, we observe syncretism between the non-specific and specific unknown markers; both are phonologically realized as *ali-*. Assuming that the non-specific marker contains only F_1, to avoid feature cumulation it is necessary that the specific unknown marker contains F_2 without F_1. This is shown in the following table and representations:

Table 19. Gapped sequence 3.

	F_1	F_2	F_3
a. *ali* (non-specific)			
b. *ali* (specific unknown)			

The trees below are based on Table 19:

(228) a. $F_1P \Rightarrow$ ali

b. $F_2P \Rightarrow$ ali

As shown above, sequence gaps may be used to form non-identical structures (sets of features) in a non-cumulative way. However, this kind of model faces certain significant problems if we consider the nanosyntactic assumptions concerning lexicalization discussed in Chapter 4. The representations shown in (226), (227) and (228) are not compatible with the most fundamental spellout principle of nanosyntax, that is the Superset Principle (as defined in Section 4.2). The challenge for the non-cumulative model stems from the fact that the Superset Principle allows a lexical entry (a lexically stored tree) to match multiple syntactic structures only if that lexical entry is a superset of those structures (syntactic trees). In other words, structures lexicalized under the Superset Principle have to be subconstituents contained within a

matching lexically stored tree. This means that a lexically stored hierarchy containing features F_1, F_2 and F_3 may match only the sets of features seen in (224), and it is impossible to lexicalize structures with gaps syncretically. In (226), for example, there can be no lexical entry that could be a superset of trees in (226-a) and (226-b) at the same time. A lexically stored tree that is a superset of (226-a) cannot be a superset of the tree in (226-b) due to the absence of F_2. Consequently, syncretism between (226-a) and (226-b) becomes impossible. The same issue arises with the distribution of features shown in (227). In a similar way, the structures in (228) also cannot be lexicalized syncretically. Again, this stems from the fact that (228-a) and (228-b) do not form a containment relation and therefore cannot be subconstituents matching a single lexically stored tree. None of the non-cumulative representations may lead to the desired forms of syncretism under the assumed formulation of the Superset Principle.

The facts presented above leave use with two options, which are either to reject the non-cumulative model or abandon the Superset Principle in the assumed formulation. I will follow the first option due to the fact that the proposed model based on feature cumulation constitutes the most efficient, straightforward and elegant way of modeling the structure of the non-specific, specific unknown and specific known indefinite markers in order to obtain the attested forms of syncretism.[63] Additionally, the model presented in Chapter 5 accurately reflects the semantic and functional compositionality of the analyzed indefinite pronouns. Since the first layer of the indefinite hierarchy can be considered to introduce only unspecified indefinite reference, the other two layers of the structure may be seen as contributing grammatical information necessary to express specific (unknown) and specific known indefinite reference. This compositionality of the indefinites is especially clear when it comes to the relation between the specific unknown and specific known indefinite functions. In order to become known (by the speaker) an indefinite referent has to be specific first, which accurately reflects the cumulative model in which

[63]Gapped structures are proposed in a number of analyses, for example Caha (2017) and Vanden Wyngaerd (2018). In these two analyses, the Superset Principle is rejected in favor of the Revised Superset Principle (Vanden Wyngaerd 2018: 11):

(i) Revised Superset Principle
"A a lexical entry L may spell out a syntactic node SN iff the features of L are equal to or a superset of the features dominated by SN."

This formulation of the Superset Principle significantly differs from the standard version of the Superset Principle. Under the standard formulation of the Superset Principle, a lexical entry matches a particular syntactic tree (a syntactic constituent) or a tree it is overspecified for. In contrast, the Revised Superset Principle allows a lexical entry to match any set of features it is overspecified for. I will not assume the Revised Superset Principle since it is incompatible with spellout-driven movement (as described in Chapter (116)).

the structure corresponding to the specific (unknown) function is syntactically contained as a subconstituent within the structure expressing the specific known function. This observation constitutes another argument in favor of a model based on feature stacking.

5.7. Analysis: summary

The purpose of the presented analysis was to show that the methodological framework of nanosyntax and its principles can be successfully employed to explain the emergence of different patterns of syncretism in paradigms of non-specific, specific unknown and specific known indefinite pronouns found cross-linguistically. The proposed model involves a syntactic hierarchy composed of three layers of syntactic features. These layers are assembled in syntax in a cumulative way and the non-specific, specific unknown and specific known indefinite functions emerge as the hierarchy grows. Each of the three indefinite marker types arises from a different subset of the indefinite hierarchy. The non-specific marker is the simplest ($[F_1]$) and the specific known marker the most complex ($[F_3[F_2[F_1]]]$), with the specific unknown marker in the middle ($[F_2[F_1]]$).

As shown in the analysis, the attested patterns of syncretism are the consequence of the indefinite hierarchy being lexicalized with different sets of lexical entries. Since, as assumed in accordance with the nanosyntactic framework, spellout of syntactic structures is regulated by the Superset Principle, a lexically stored syntactic tree can match multiple sets of syntactic features as long as those features form structures contained within one another and the lexically stored tree constitutes their superset (proper or not). Hence, in the presented model, the three possible patterns of syncretism (AAA, AAB and ABB) emerge when one or two lexically stored trees are matched with the indefintie hierarchy. The AAA pattern appears when only one lexical entry is available for the lexicalization of the indefinite hierarchy, while the AAB and ABB patterns are the consequence of two lexical entries being available in a language. In cases where three lexical entries are used to lexicalize the hierarchy, each indefinite marker type receives a different phonological exponent (ABC). Additionally, the proposed model accurately rules out the possibility of the unattested ABA pattern due to the application of the Elsewhere Principle which guarantees that the most specific lexical entry is chosen in every spellout cycle.

The spellout algorithm put forward as a part of the nanosyntactic framework applied to the proposed indefinite hierarchy properly explains not only the attested patterns of syncretism but also the morphological distribution of phonological exponents of the indefinite markers. The

nanosyntactic mechanisms of spellout-driven movement and subderivation allow us to show a clear syntactic different between structures that phonologically surface as prefixes and those that are lexicalized as suffixes. Hence, it becomes possible to create a model of derivation for indefinite markers (phonological exponents) which accounts for their positioning with respect to indefinite pronoun bases. Indefinite pronouns derived through spellout-driven movement will surface as suffixes (with a unary foot), while the mechanism of subderivation will give rise to prefixes (with a binary foot).

Lastly, the nanosyntactic model of spellout can also be used do deal with cases in which indefintie pronouns do not have the most common base + marker form but instead are monomorphemic or syncretic with wh-pronouns. Since phonological exponents are inserted into phrasal nodes, a single exponent may lexicalize the whole structure of an indefinite pronoun. Additionally, as permitted by the Superset Principle, the same phonological exponent may be inserted into multiple adjacent phrasal nodes, which explains how different types of monomorphemic indefinites may be syncretics with one another and their bases. According to the presented mmodel, the lexical exponent inserted into a phrase, for example a base, is overwritten by the same exponent when more features are added to the structure. In consequence, the lexicalization outcome remains the same for multiple non-identical structures.

Chapter 6: Loose ends

6.1. Unproductive markers

This section discusses indefinite marker variation and explores the mechanism of pointers as a solution to the issue of irregular and unproductive morphology. First, consider the data from Bulgarian where the assertive indefinite series of pronouns is formed by the prefix *nja-*. The only exception in the paradigm is *ne-što* 'something', which bears the marker *ne-*:

(229) Bulgarian (Haspelmath 1997: 267)

a. *ne*-što
INDEF what
'something'

b. *nja- koj*
INDEF who
'someone'

c. *nja- kade*
INDEF where
'somewhere'

d. *nja- koga*
INDEF when
'sometime'

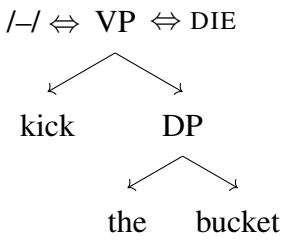
e. *nja- kak*
INDEF how
'somehow'

f. *nja-* *kolko*
INDEF how.much
'some amount'

The indefinite marker *ne-* does not differ in meaning or function from the standard prefix *nja-*, but it is not a productive morpheme. It appears only in combination with one particular base, namely *što* 'what'.[64] If there is no functional or semantic difference between *nja-* and *ne-*, we can conclude that both morphemes are structurally uniform (they realize the same syntactic structure). What remains a question then is why the indefinite prefix receives a special phonological exponent when merged on top of *što*.

A solution in line with the nanosyntactic system of lexicalization is a mechanism known as pointers (De Clercq and Vanden Wyngaerd 2019). A pointer is a node inside a lexical entry that refers to (points to) a lexical item. For example:

(230) Pointers



In the lexical entry shown in (230), the pointers (marked with arrows) point to lexical items stored in the lexicon. Therefore, this entry will match a structure which is spelled out as these particular lexical items, that is *kick the bucket*. When spellout happens at the VP node, the semantic value associated with the entry in (230) will overwrite the previous meaning of the structure inside *kick the bucket*. At the same time, no new phonology will be inserted due to the fact that none is specified by the lexical entry. The final outcome is an idiom which connects the phonological value of *kick the bucket* with the meaning of DIE.

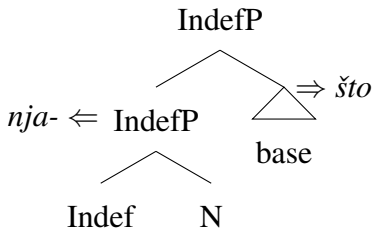
As argued in De Clercq and Vanden Wyngaerd (2019), pointers can be used not only to explain idiomatic expressions but also to tackle the issue of unproductive morphemes. The indefinite marker *ne-* in Bulgarian ((229)) can be considered an unproductive morpheme since it is used to form only one item in its paradigm, as opposed to the regular marker *nja-*. Below, I

[64] *Što* is also not a standard form. in modern Bulgarian, it has been replaced by *kakvo* 'what'.

will show how we can use the mechanism of pointers to explain the existence of the *ne-* marker.

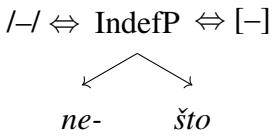
First, assume that the derivation of the indefinite hierarchy on top of the base *što* follows the same steps that take place when the indefinite features are merged with other bases in the paradigm. Since there is no lexical entry that can be used to spell out F_1 and the rest of the hierarchy together with the base, the lexicalization system will alter the geometry of the structure in accordance with the spellout algorithm. When the final last-resort option is reached, a prefix will be formed through subderivation. This prefix will be lexicalized with the regular phonological exponent *nja-*. This gives us the following representation:

(231)



Of course, this is not the desirable result since the target form is *ne-što*, not *nja-što*. To solve this problem and turn *nja-* into *ne-*, we can propose that the lexicon contains the following lexical entry:

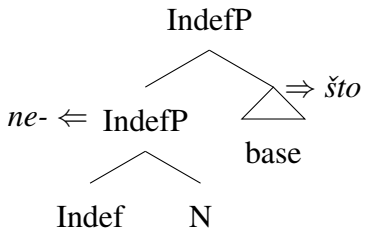
(232) Lexical entry



This lexical entry does not introduce any new phonology (/–/) or semantic information ([–]) at the IndefP node and points to two lexical items, that is an pronominal base *što* and an indefinite marker *ne-*. When the subderived indefinite prefix (IndefP) is merged with the base, the spellout system will consult the lexicon in order to see if there is a lexical entry which can match the newly formed topmost node (IndefP). In line with De Clercq an Vanden Wyngaerd (2019), I propose that at this point, the lexical entry in (235) will become a match for the node and the unproductive marker *ne-* will overwrite the regular marker *nja-*. The unproductive lexical item *ne-* in (235) can match the indefinite marker *nja-* in (231) due to the fact that it is present in the lexicon and corresponds to the same structure as the standard prefix *nja-* ([IndefP [Indef] [N]]). However, *ne-* is unavailable for insertion through normal means and can be accessed only via a

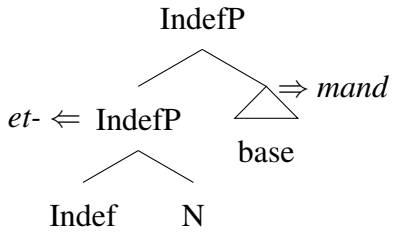
pointer. It is the role of the pointer in (235) to specify in what circumstances *ne-* should be made available for insertion. Thus, the lexical exponent *ne-* will be made available and overwrite the standard phonological exponent of the indefinite prefix:

(233)



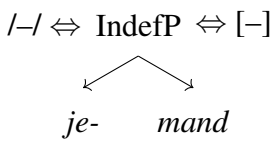
An analysis based on the idea of pointers can also be applied to other cases in which we observe indefinite marker variation for the same function and meaning. Another instance of this kind of variation are the German *et-was* 'something' and *je-mand* 'someone'. These two indefinites appear in the same indefinite functions (non-specific, specif unknown and specific known) but bear phonologically different markers. It can be argued that in this case a pointer will also specify which indefinite marker exponent should (*ja-*/*et-*) appear with which base (*mand*/*was*). Since *et-was* and *je-mand* are the only forms in their series, assume that *et-* is the standard exponent for the indefinite prefix. When merged with the indefinite base *-mand*, the prefix will initially have *et-* as its phonological exponent:

(234)



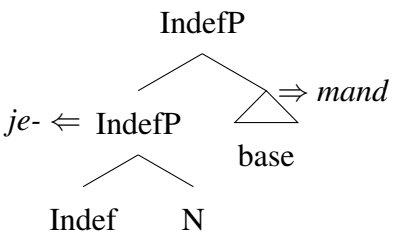
The phonological form of the prefix will subsequently be overwritten by *je-*, which will be available for insertion by the following lexical entry:

(235)



Thus the mechanism of pointers allows us to obtain the right lexicalization for German *je-mand* 'someone':

(236)



The system of pointers allows us to account for the existence or unproductive morphology in a straightforward way without postulating structural differences between regular and irregular morphemes. Additionally, as shown above, the mechanism of pointer can be used not only in the case of adjectives, discussed in De Clercq an Vanden Wyngaerd (2019), but also to explain morphological variation in other domain of grammar.

6.2. Paradigm gaps

The cross-linguistic comparative study conducted the non-specific, specific unknown and specific known indefinite markers revealed a number of languages in which indefinites corresponding to one or more of the studied indefinite functions are absent.[65] This means the lack of a pronoun series that could be used to express a particular indefinite function. Table 20 shows languages in which gaps of this kind have been attested:

Table 20: Indefinite markers. paradigm gaps.

	non-specific	specific unknown	specific known
Kannada	-aadaruu	-oo	–
Veps	-ni	-ni	–
Quachua	-pis/-pas	-chi/-cha	–
Chinese	wh-pronoun	–	–
Swahili	–	–	–
Irish	–	–	–
Filipino	–	–	–

The absence of one or more indefinite pronouns series in a language can be argued to be connected with the lack of indefinite markers expressing particular indefinite functions. If

[65] See Section 3.6 for data concerning languages with paradigm gaps.

this is the case, then the explanation for the data in Table 20 will be connected with how the indefinite hierarchy lexicalized in a particular language.

Following the assumption that the functional sequence is cross-linguistically universal (see Section 4.1), we can reduce the problem of paradigm gaps to language-specific contents of the lexicon. A language will not have a marker and in consequence an indefinite pronoun series for a particular indefinite function if its lexicon does not contain a lexical entry which could be used to lexicalize that marker. Thus, even if a set of features corresponding to an indefinite function can be projected, it will not be realized as an indefinite marker and used to form an indefinite pronoun series. Also, consider the fact that the absence of an indefinite marker type in a given language does not mean that its function cannot be expressed in a different way. We see this, for example, in Irish and Swahili, which both have indefinite modifiers corresponding to the assertive indefinite functions.[66] Speakers of Irish may use the modifier *éigin* 'some/certain' to express different kind of indefinite meaning:

(237) Bhaldraithe (1959: 683)

a. *Dúirt* ***duine éigin*** *liom é.*
told person certain to:me he
'Some person (specific) told me.'

b. *Abair* ***rud éigin****.*
say:IMPV thing certain
'Say some thing (non-specific).'

The modifier *fulani* 'some' is used in a similar way in Swahili:

(238) Mpiranya (2015: 87, 203)

a. *Ni muhimu* ***mtu fulani*** *asaidie watoto.*
'It is important that some person (non-specific) help the children.'

b. *Wahuni wamemwibia* ***mtu fulani*** *saa yake.*
'Hooligans have robbed some person (specific) of his watch'

The idea that indefinite marker type gaps can be reduced to how a given language can

[66]It is not the case that Irish, Filipino and Swahili completely lack any indefinite pronoun forms. For example, Filipino and Swahili have free choice pronoun series. In Irish, there seem to be very few indefinite forms; one of them is *dada* (anything).

lexicalize the indefinite hierarchy is supported by a generalization that may be formulated on the basis of the data shown in Table 24. Although the number of languages with indefinite pronoun type gaps in the studied language sample is rather small, the data collected so far (Table 20) suggests that such gaps follow a particular pattern. If only one indefinite pronoun type is missing, it will be the specific known type which requires the whole indefinite structure to be assembled. Pronouns of the specific unknown type may be absent if specific known indefinites are absent as well. Lastly, non-specific indefinite pronouns are not present if the other two types are missing as well. If this generalization is correct, the regularity can be argued to stem directly from one of the main principles of lexicalization in nanosyntax, namely the Superset Principle. As discussed in Chapter (116), under the Superset Principle, if a lexically stored tree matches a syntactic structure, it can also be used to lexicalize any subsets of that structure (subconstituents). In the case of indefinite markers, this will mean that a marker will never remain without a phonological exponent if it is possible to spell out markers which constitute its supersets.[67]. Hence, we can predict that an indefinite marker corresponding to a particular function has to be present in a language if that language has a marker representing an indefinite function which is higher in the paradigm (a more complex function). This is exactly what we see in Table 20.

Linguistic variation reduced to the contents of the lexicon may also explain another kind of gaps, that is indefinite pronoun base gaps. First, consider the assertive indefinite pronoun paradigm in English. The following table shows what bases are used to form English assertive indefinite pronouns:[68]

Table 21. Indefinite pronoun paradigm in English.

category	wh-pronoun base	nominal base	*one*-pronoun base
THING	–	some-thing	–
PERSON	–	some-body	some-one
PLACE	some-where	some-place	–
MANNER	some-how	some-way*	–
AMOUNT	–	–	–
TIME	some-when*	some-time	–

[67]See, for example, Section 5.7.

[68]Forms marked with * are rare but attested. The Corpus of Contemporary American English shows 283 example sentences with *someway* and 19 sentences with *somewhen*.

As shown in Table 21, English has developed a paradigm where all three types of bases are used. However, indefinite pronouns for certain base types and ontological categories are not attested, for example *some-who*. Note that this indefinite form is absent from the paradigm even though it could theoretically be allowed since *who* exists.

A solution to this issue can be proposed but it requires us to assume a structural difference between actual bases and forms that could become bases. This means a feature (or features) preceding the indefinite hierarchy in the fseq and lexicalized together with the base. In consequence, indefinite pronoun bases become limited only to forms that can lexicalize that feature. In cases where it is impossible to lexicalize the proposed feature (e.g. English *who*), the merger of the indefinite hierarchy will not proceed.

It can be argued that the existence of such a structural difference is suggested by special bases which are bound morphemes and may appear only together with indefinite markers.[69] This kind of bases take the form of neither nouns or pronouns that can be stand-alone lexical items in a given language:

(239) Special bases (in bold)[70]

a. *je-**mand*** (German: 'somebody')
b. *ie-**ts*** (Dutch: 'something')
c. *xi-**mkien*** (Maltese: 'somewhere')
d. *någon-**stans*** (Swedish: 'somewhere')
e. *ali-**cubi*** (Latin: 'somewhere')
f. *bir-**närse*** (Kazakh: 'something')
g. *ne-**što*** (Bulgarian: 'something')

It can be suggested that special bases will arise when the proposed additional feature is not lexicalized syncretically with a functional noun or interrogative/relative pronoun. If there is

[69]However, it has to be noted that the existence of such special bases can also be explained with the use of pointers. This means that special bases could be similar to unproductive indefinite markers. See Section 6.1.

[70]The bases in (239), are not stand-alone forms and appear only in combination with indefinite markers. The German base *-mand* is related to Mann 'man'. The Dutch *-ts* comes from an earlier form *wiht* 'thing'. In Maltese, the form *mkien* is related to *makaan* 'place'. The Sweedish *-stans* can be traced back to Old Norse *staðr* 'place/location' and it does not seem to be related to the modern Swedish *plats* 'place'. In Latin, *-cubi* in *ali-cubi* 'somewhere' is connected with an earlier form of *ubi*, namely *quobi* 'where'). As for *što* 'what', it seems to have been replaced by *kakvo* 'what' (related to *kakav* 'what kind of') in modern Bulgarian. Lastly, the origin of *-närse* in Kazakh *bir-närse* 'something' is unknown to me (Haspelmath 1997: 244, 246, 249, 253, 288, 293, Lewis 1891).

no lexical entry that can lexicalize a particular base, the derivation will not continue and the indefinite hierarchy will not be merged. In consequence, a paradigm gap is created. However, this solution should be treated as a tentative proposal, rather than a definite answer to the issue of indefinite base gaps.

6.3. Overlapping forms

In a number of languages, the non-specific, specific unknown and specific known functions are expressed by multiple indefinite markers (phonological exponents) simultaneously. Examples of this can be seen in languages such as German, Czech, Slovak, Russian, Polish and Greek. In each of these languages, there are additional indefinite pronoun series that correspond to the assertive indefinite functions. Below, I analyze the examples of overlapping forms found in the studied language sample and discuss possible explanations.

6.3.1. German: *irgend-* series

First, consider the data from German, where the standard indefinite pronoun series that covers the non-specific, specific unknown and specific known functions contains only two forms, namely *et-was* 'something' and *je-mand* 'someone'. Other ontological categories such as PLACE, TIME and MANNER have to be expressed by wh-based pronouns belonging to the *irgend-* series. This series of indefinites may appear in the specific unknown, non-specific and free choice functions (cf. Aloni and Port 2013):

(240) German

a. *et-* *was*
INDEF what
'something'

b. *je-* *mand*
INDEF PERSON
'something'

(241) German *irgend-* indefinites

a. *irgend- was*
INDEF what
'something/anything'

b. *irgend- wer*
INDEF who
'someone/anyone'

c. *irgend- wo*
INDEF where
'somewhere/anywhere'

d. *irgend- wann*
INDEF when
'sometime/anytime'

e. *irgend- wie*
INDEF how
'somehow/anyhow'

The matter becomes more complicated if we take into account a few other facts about the German indefinite pronoun system. Apart from the wh-based forms *irgend-wer* and *irgend-was*, the *irgend-* series also contains indefinite pronouns for the PERSON and THING categories which are derived from the basic *etwas* and *jemand* indefinites, namely *irgend-jemand* and *irgend-etwas*:

(242) German *irgend-* indefinites

a. *irgend- et- was*
INDEF INDEF what
'something/anything'

b. *irgend- je- mand*
INDEF INDEF PERSON
'someone/anyone'

In colloquial speech, speakers may also use indefinites syncretic with wh-pronouns, for example

wer 'who/someone' and *was* 'what/something':[71]

(243) German indefinite pronouns syncretic with wh-pronouns

Da ist ***wer*** *an der Tür.*
there is who by the door
'There is **somebody** at the door.'

Now compare the functional scope of each of the series mentioned above. The *jemand/etwas* series may appear in the non-specific, specific unknown and specific known contexts. The same seems to be possible for the wh-indefinites:

(244) a. *Ich habe* ***was/etwas*** *verloren. Rate mal, was?*
I have something lost guess PART what
'I lost **something** (specific known). Guess what?'

b. *Da ist* ***wer/jemand*** *an der Tür.*
there is someone at the door
'There is someone at the door.'

c. *Kannst du mir* ***was/etwas*** *mitbringen?*
can you me something bring
'Can you bring me **something**?'

In contrast, indefinites in the *irgend-* series are used in the non-specific and free choice functions. They may also be interpreted as specific unknown is some cases:[72]

(245) a. *Auf was für einen Film hast du Lust? Hmmm keine Ahnung...*
on what for a film have you desire hmm no idea
irgend-was *lustiges.*
anything/something funny
'What kind of movie do you feel like watching? Hmmm ... no idea... **anything** funny/**something** (non-specific) funny.'

[71] Dutch also allows wh-indefinites in colloquial speech. Since I have little data on Dutch wh-indefinites, I do not discuss this language separately. For some examples, see Hengeveld et al. (2021).

[72] A speaker I interviewed showed a tendency to interpret *irgend-* indefinites as either non-specific or free-choice items.

b. *Da ist* ***irgend-jemand*** *an der Tür.*
there is someone at the door
'There is someone at the door.'

c. *Dieses Problem kann* ***irgend-jemand*** *lösen.*
this problem can anyone solve
'**Anyone** can solve this problem.'

An analysis of indefinite pronouns syncretic with interrogative pronouns in presented in Section 5.5. In line with that proposal, I propose that the observed syncretism is the result of the indefinite hierarchy being lexicalized together with pronominal bases. Speakers will access different lexical entries depending on the social context, and in consequence use either standard on interrogative pronoun forms:

(246) a. [Indef [*wer*]] ⇒ *wer* (indefinite)
b. [Indef [*was*]] ⇒ *was* (indefinite)

In the case of *irgend-* indefinites, they appear to constitute an emphatic series that is used to stress the ignorance or indifference of the speaker. The fact that the *irgend-* prefix may be added to items which already are indefinite pronouns by themselves (e.g. *jemand* and *etwas*) suggests that it contains features that are not present in the assertive indefinite hierarchy. This conclusion is also supported by the fact that *irgend-* indefinites may express the free choice function, which is impossible for specific/non-specific pronouns:

6.3.2. Polish *-ś* and *nie-* forms

The general specific/non-specific indefinite pronoun series in Polish is formed by adding the *-ś* morpheme to pronominal bases. Consider some basic examples:

(247) Polish

a. ***kto*** *-ś*
who INDEF
'someone'

b. ***co*** *-ś*

what INDEF

'something'

c. ***gdzie*** *-ś*

where INDEF

'somewhere'

d. ***kiedy*** *-ś*

when INDEF

'sometime'

e. ***jako*** *-ś*

how INDEF

'somehow'

f. ***ile*** *-ś*

how.much INDEF

'some amount'

The *-ś* series of indefinites appears to overlap with the *nie-* series of indefinites. This indefinite series however has not retained its original meaning and only *nie-który* and perhaps *nie-kiedy* are used as a regular indefinite forms:

(248) Polish

a. *nie-* *który*

INDEF which

'some (but not all)'

b. *nie-* *gdzie*

INDEF where

survives only in the fixed phrase *gdzie-nie-gdzie* 'in some places'

c. *nie-* *kiedy*

INDEF when

'sometimes'

d. *nie-* *jak-i*
INDEF how-ADJ
'certain/some kind of'

e. *nie-* *jako*
INDEF how
'in a way/somewhat/as if'

f. *nie-* *co*
INDEF what
'a bit/somewhat'

The indefinite marker *nie-* expresses a rther specialized meaning which does not overlap with the semantics of the regular indefinite marker *-ś*. Forms such as *nie-który* 'some (but not all)' and *nie-kiedy* 'sometimes' indicate that *nie-* can be interpreted as 'some of a set but not all'. This indicates that *nie-* realizes a different set of features than *-ś*.

The other forms in the *nie-* series appear to be idiomatic since their meanings do not follow directly from their morphological composition. Perhaps, the meaning of *nie-co*, *nie-jaki*, *nie-jako* and *gdzie-nie-gdzie* could be explained with the use of pointers (see Section 6.1). As argued in De Clercq and Vanden Wyngaerd (2019), the mechanism of pointers can overwrite the structural semantics of a phrase without changing its phonological form. This can be show on the basis of *nie-jaki* 'certain/some kind of':

(249) Polish

Powiedziała to z nie-jakim zażenowaniem.
said.3SG this with INDEF-how-ADJ embarrassment
'She said it with some/certain embarrassment.'

The compositional meaning of *nie-jaki* can be overwritten by the following lexical entry:

(250) Lexical entry

/–/ ⇔ IndefP ⇔ CERTAIN/SOME KIND OF

nie- *jaki*

When the form *nie-jaki* is formed, it will match the lexical entry in (250) and have its original meaning overwritten. In this way we can use the technology of pointers to explain the non-compositional meaning of grammaticalized forms.

6.3.3. Russian: *-libo* and *ne-* indefinites

In Russian, the indefinite hierarchy is realized as three indefinite pronoun series, namely the non-specific *-nibud*, specific unknown *-to* and specific known *koe-* indefinites. Apart form these three, there is also the formal non-specific *-libo* series and a marginal *ne-* series:[73]

(251) Russian

a. *što -nibud* / *što -libo*
what INDEF / what INDEF
'something' (non-specific)

b. *što -to* / *ne- što*
what INDEF / INDEF what
'something' (specific unknown)

c. *koe- što*
INDEF what
'something' (specific known)

As claimed in Eremina (2012: 72) and confirmed by a native speaker, there is no difference in meaning between *-nibud* and *-libo* forms. The *libo* series appears as a formal variant of the regular *-nibud* series, usually in written language. It can be argued that the *-nibud*/*-libo* variation is a case in which speakers have two lexical entries available for a single syntactic structure (the non-specific marker). It is up to the speaker which phonological exponent will be inserted in a given context (formal/informal).

Items in the *ne-* series no longer seem to have consistent semantics. Forms *ne-kto* 'someone' and *ne-što* 'something' are not regular indefinites anymore and appear only in nominative and accusative cases. They express the specific unknown function and are often used emphatically to describe things or beings that evoke emotions such as surprise of fear. The remaining items in the series, except *nekiy* 'certain' and *nekogda* 'once/formerly' seem to have

[73] All forms have pronominal bases.

negative meaning:[74]

(252) Russian

a. *ne-* *kto*
INDEF who
'somebody'

b. *ne-* *čto*
INDEF what
'something'

c. *ne-* *gde*
INDEF where
'nowhere'

d. NE- KOGO
INDEF who.ACC
'nobody.ACC'

e. *ne-* *čego*
INDEF what.ACC
'nothing.ACC'

f. *ne-* *kogda*
INDEF when
'once/formerly'

The *ne-* indefinites are not the standard way of expressing indefinite meaning in Russian. Forms such as *ne-kto* 'someone' and *ne-što* 'something' are used emphatically, which suggests structural differences between *ne-* and the regular indefinite markers. The form *ne-kogda* 'once/formerly' appears to have become idiomatic (see Section 6.3.2). Lastly, ne-kogo, ne-čego and other similar forms behave like n-words. Due to the complexity of this case, I will not discuss it further. It would be necessary to conduct a separate detailed study of *ne-* forms to properly understand their use and how exactly they differ form the regular *-nibud*, *-to* and *koe-* indefinites.

[74]I provide only accusative forms for the negative indefinites.

6.3.4. Czech *-si* and Slovak *-si* and *vol'a-*

Czech is a language that belongs to the AAA category. The main indefinite pronoun series is marked with the prefix *ně-* and realizes all three indefinite functions (non-specific, specific unknown and specific known). There is however another indefinite pronoun series with the suffix *-si*, which seems to be allowed only in the specific unknown function:

(253) Czech

a. *ně-* *co*
INDEF what
'something' (non-specific)

b. *ně-* *co* / *co* *-si*
INDEF what / what INDEF
'something' (specific unknown)

c. *ně-* *co*
INDEF what
'something' (specific known)

Lukeš and Kauerová (2012: 34) remarks that *-si* can be considered an archaic version of *ně-* and some items in the series no longer retain their original meaning. For example, *kdy-si* (when-INDEF) is used in the meaning of 'a long time ago' instead of 'sometime' (as implied by the form of the pronoun). Forms that have not undergone a semantic shift correspond solely to the specific unknown function, which indicates that the indefinite marker *-si* may be structurally similar to the specific unknown *ně-*. Presumably, *-si* indefinites contain the first two layers of the indefinite hierarchy and some other feature sequence merged above or below them.[75]

In Slovak, the three indefinite functions are covered by at least three indefinite pronoun series, namely the general *nie-* series, the *-si* series and the *vol'a-* series. The first series of pronouns (*nie-*) covers all three indefinite functions, while the other two series (*-si* and *vol'a-*) may appear only the non-specific and specific unknown functions:

[75] As mantioned by a native speaker, some speakers seem to use *-si* forms the same way the *ně-* series is used. Perhaps, for those speakers *-si* is just a phonological variant of *ně-*.

(254) Slovak

a. *nie- čo / vol'a- čo / čo -si*
INDEF what / INDEF what / what INDEF
'something' (non-specific)

b. *nie- čo / vol'a- čo / čo -si*
INDEF what / INDEF what / what INDEF
'something' (specific unknown)

c. *nie- čo*
INDEF what
'something' (specific known)

As suggested by a corpus study presented in Richtarcikova (2013), these two indefinite series appear mainly in the specific unknown function. Other uses were also attested, for example the non-specific function, but were significantly less frequent. This makes the *-si* and *vol'a-* series similar to Czech *-si* indefinites which are specific unknown forms.[76]

6.3.5. Greek

In Greek, the general *ka-* series of pronouns may be replaced by the *tipota/kanenas* (something/someone) series in the non-specific function (Haspelmath 1997: 266 and Holton et al.):

(255) Greek

a. *ka- ti / ti -pota*
INDEF what / what INDEF
'something' (non-specific)

b. *ka- ti*
INDEF what
'something' (specific unknown)

c. *ka- ti*
INDEF what
'something' (specific known)

[76]There seems to be a preference among speakers to use *vol'a-* forms in the non-specific function more frequently than *-si* forms (Richtarcikova 2013).

The *tipota/kanenas* series is used as an NPI series in questions and negative statements. Indefinites of this kind are also not compatible with positive declarative sentences:[77]

(256) Giannakidou (2001: 3), Haspelmath (1997: 266)

a. **Idha* ***kanenan***.
saw.PERF.1SG INDEF.PERSON
'*I saw **anybody**.'

b. *Dhen idha* ***kanenan***.
not saw.perf.1sg INDEF.PERSON
'I didn't see **anybody**.'

c. *Idhes* ***ti-pota*** **/** ***ka-ti***?
saw.2SG what-INDEF / INDEF-what
'Did you see **anything/something**?'

Giannakidou (2001), Giannakidou (2011b) and Holton et al. (2004: 96-98) indicate that indefinite pronouns of the *tipota/kanenas* series are not non-specific indefinites. As negative/affective polarity items, these pronouns are licensed not only by negative and question environments but also by non-veridical contexts such as question, conditional, subjunctive and imperative clauses. This makes pronouns such as *tipota* or *kanenas* appear in sentences where the referent is non-specific together with regular non-specific indefinites of the *ka-* series.

[77] The *tipota/kanenas* indefinites may also function as n-words (without a sentential negator).

Summary and conclusions

The main aim of this dissertation was to propose a representation of the syntactic structure of assertive indefinite markers in which the particular types of these markers realized different subsets of a single, cross-linguistically universal hierarchy of syntactic heads. On the basis of a cross-linguistic comparative study it has been argued that assertive indefinite pronouns can be divided into non-specific, specific unknown and specific known types. Indefinite pronouns used in the non-specific function appear in contexts that do not involve any particular individual referent or when the speaker is ignorant or indifferent as far as the existence of a referent is concerned. Specific indefinites differ from non-specific forms as they always have referents that the speaker can identify as particular entities. The specific unknown type refers to individual entities whose exact identity is not known to the speaker, while specific known indefinite pronouns have referents that are particular entities which the speaker can identify.

As suggested in the implicational map of indefinite functions put forward in Haspelmath (1997), the non-specific, specific unknown and specific known indefinite functions can be arranged in a particular order which reflects how these functions may be expressed by indefinite pronoun series cross-linguistically. The relative order of these functions, as indicated in the implicational map, was confirmed by the analysis of syncretism between indefinite markers corresponding to the non-specific, specific unknown and specific known indefinite functions. As shown, the non-specific, specific unknown and specific known indefinite markers form a paradigm in which we observe three patterns of syncretism with the simultaneous absence of the ABA pattern. This means that the proposed paradigm is in line with the predictions of the *ABA generalization proposed in Bobaljik (2007, 2012), according to which the ABA pattern of syncretism should not arise in an ordered set of forms:

Table 22. Indefinite pronoun paradigm: patterns of syncretism.

pattern	non-specific	specific unknown	specific known
AAA	■	■	■
ABB		■	■
AAB	■	■	
*ABA	■		■

The paradigm of the non-specific, specific unknown and specific known indefinite markers was analyzed in accordance with the nanosyntactic model of grammar and its core tenets such as the Superset Principle, the Elsewhere Principle and phrasal spellout. As argued in Chapter 5, the attested patterns of syncretism and the absence of the ABA pattern reveal the shared syntactic makeup of the three types of assertive indefinite markers, which can be represented as a syntactic hierarchy composed of three layers of structure. Thus, it has been claimed that the particular marker types result from stacking syntactic features in accordance with a particular sequence. In other words, non-specific, specific unknown and specific known indefinite markers constitute phonological realizations of three sets of features based on the hierarchy:

(257) a. $[F_1] \Rightarrow$ non-specific marker

b. $[F_2\ [F_1]] \Rightarrow$ specific unknown marker

c. $[F_3\ [F_2\ [F_1]]] \Rightarrow$ specific known marker

While syncretism shows only the relative order of elements in the hierarchy, the simplest and the most structurally complex marker is established on the basis of their functional semantics. As the functional compositionality indicates, the non-specific marker is the simplest, and the specific known marker contains the whole indefinite hierarchy. The specific unknown marker corresponds to only two layers of the structure. This kind of representation is consistent with the proposed containment structure and reflects the change of functional meaning from non-specific to specific and then from unknown to known.

The indefinite hierarchy can be projected on top of three types of bases, namely functional nouns, wh-pronouns and generic *one* pronouns. What makes these forms indefinite pronoun bases is the presence of the functional ontological categories in their structure. These ontological categories will determines the type of referent expressed by a particular base.

Subsequently, it has been shown that the proposed indefinite hierarchy allows us to derive all the attested patterns of syncretism through phrasal spellout. The form of syncretism in a given language depends on the number of lexical entries that can match the three layers of the indefinite hierarchy. When only one phonological exponent realizes the three constituents formed by the layers of the hierarchy, the AAA pattern (full syncretism) arises. If there are two lexical entries that match the hierarchy, either the AAB or the ABB pattern may be created. Lastly, in cases where each subset of the hierarchy is matched by a different lexically stored tree, we observe no syncretism (ABC). As argued in Section 5.6, these results cannot be obtained in a model which does not involve derivation through feature cumulation. As shown, under the standard formulation of the Superset Principle (cf. Starke 2009), syncretim may only emerge as a result of syntactic containment. In other words, the Superset Principle entails that syncretism is bound to target only adjacent phrasal nodes in a hierarchy.

The nanosyntactic system of lexicalization and its spellout algorithm is also used to explain the morphological distribution of indefinite markers with respect to their bases. If at any stage of the derivation, a matching lexical entry cannot be found, the spellout system will follow a sequence of last-resort operations until a lexicalizable tree geometry can be obtained. According to the discussed spellout algorithm, suffixal indefinite markers are formed through spellout-driven movement, which involves displacement of a piece of structure to a position above the newly merged feature. In consequence, the remnant constituent (with a unary bottom) is lexicalized as a suffix with respect to the displaced phrase. In contrast, prefixes are created as a result of the final last-resort operation, which is the mechanism of subderivation. The process involves the formation of a separate derivation which is later integrated into the main structure as a left-branch constituent (with a binary bottom). Thus, not only does the nanosyntactic system of lexicalization provide an explanation for the morphological positioning of indefinite markers but also draws a clear line between suffixes and prefixes since these two types of structures now differ with respect to their syntactic geometry (unary vs. binary bottom).

The proposed representation of the internal structure of non-specific, specific unknown and specific known indefinite pronouns and their corresponding markers, and the subsequent analysis based on the principles of the nanosyntactic framework successfully explain the collected data. As it has been argued, it is possible to show how syncretism and the *ABA generalization indicate that assertive indefinite pronouns constitute another domain of grammar with a complex internal structure in the form of a cross-linguistically universal hierarchy. In the future,

it may be possible to extend this analysis to other types of indefinite pronouns and indefinite functions. As suggested by the implicational map of indefinite functions presented in Haspelmath (1997), NPI, free choice and negative pronouns may be in some way, perhaps structurally, related to the assertive indefinite category. However, to answer this question, further comparative studies are necessary.

Abstract

This dissertation argues that indefinite markers used to form the non-specific, specific unknown and specific known types of assertive indefinite pronouns lexically realize a cross-linguistically universal hierarchy of syntactic heads. As revealed by a cross-linguistic comparative study, assertive indefinite pronouns can be divided into non-specific, specific unknown and specific known on the basis of their functional semantics. The same analysis of data also demonstrates that in a given language, these three functional types can be expressed by one (as in English), two (as in Georgian) or as many as three different indefinite pronoun series (as in Russian). These facts are taken as evidence for indefinite pronoun, and in consequence also indefinite marker, syncretism.

The paradigm of the non-specific, specific unknown and specific known indefinite marker types established on the basis of the collected data reveals clear patterns of syncretism. Additionally, the paradigm demonstrates assertive indefinite pronouns to be another domain that conforms with the *ABA generalization (Bobaljik 2007, 2012) since the ABA pattern remains unattested.

As argued, syncretism between the three types of indefinite markers and the lack of the ABA pattern indicate the existence of an underlying hierarchy composed of three layers of syntactic features. The elements of the hierarchy are merged in a cumulative way (stacked), which means that the particular indefinite marker types emerge from the hierarchy in a sequence as the structure grows. In other words, non-specific, specific unknown and specific known indefinite markers are syntactically contained within one another since they realize three different subsets of the indefinite hierarchy.

The proposed representation allows us to explain the cross-linguistically attested patterns of syncretism through the application of the methodological principles of nanosyntax (Starke 2009, 2011) such as the Superset Principle, the Elsewhere Principle and phrasal spellout. The observed forms of syncretism arise when a single phonological exponent spells out mul-

tiple layers (constituents) of the indefinite hierarchy, as permitted under the Superset Principle. At the same time, the absence of the ABA pattern is explained through the Elsewhere Principle, which guarantees that syncretism may target only adjacent layers of the hierarchy. Thus, each of the attested patterns of syncretism can be shown to stem from a different number of phonological exponents that are used to lexicalize the indefinite hierarchy in a given language. As a result, the analysis shows that assertive indefinite pronouns constitute another domain of grammar in which syncretism and the *ABA generalization reveal the complex internal syntactic structure.

Streszczenie

Niniejsza rozprawa doktorska przedstawia morfoskładniową analizę twierdzących zaimków nieokreślonych (np. *coś* i *ktoś*). Według przedstawionej analizy afiksy używane do tworzenia trzech typów twierdzących zaimków nieokreślonych stanowią leksykalną realizację uniwersalnej hierarchii cech gramatycznych. Badanie porównawcze przeprowadzone dla potrzeb tej rozprawy pokazuje, że twierdzące zaimki nieokreślone mogą być podzielone na zaimki, które nie mają indywidualnego odniesienia (non-specific), zaimki, które mają konkretne indywidualne ale nieznane dla mówiącego odniesienie (specific unknown) i takie, których odniesienie jest znane mówiącemu (specific known). Te trzy rodzaje zaimków nieokreślonych mogą być fonologicznie wyrażone za pomocą jednej serii form (np. w języku polskim), dwóch (np. w języku gruzińskim) lub nawet trzech (np. w języku rosyjskim). Na podstawie tych obserwacji można argumentować, że istnieją trzy funkcjonalnie różne typy twierdzących afiksów nieokreślonych, pomiędzy którymi może występować synkretyzm (fonologiczna identyczność).

Stwierdzone w badaniu porównawczym formy synkretyzmu afiksów nieokreślonych pozwalają na ustalenie ich paradygmatu. Kolejność form w tym paradygmacie ujawnia różne formy synkretyzmu z wyjątkiem wzoru ABA. Oznacza to, że twierdzące afiksy nieokreślone stanowią kolejną kategorię gramatyczną, która jest zgodna z generalizacją *ABA (Bobaljik 2007, 2012). Odkryte formy fonologicznej identyczność trzech typów twierdzących afiksów nieokreślonych, a także brak synkretyzmu typu ABA, wskazują na istnienie uniwersalnej składniowej hierarchii cech gramatycznych odpowiadającej tym afiksom. Zanalizowane typy afiksów nieokreślonych (non-specific, specific unknown i specific known) są formowane składniowo w konkretnej kolejności poprzez konstruowanie proponowanej hierarchii w sposób kumulatywny. W ten sposób strukturalnie mniejsze afiksy są składniowo zawarte w strukturze bardziej złożonych form, a każdy typ afiksu zawiera inną liczbę cech gramatycznych opartą na proponowanej hierarchii.

Przedstawiona składniowa reprezentacja struktury twierdzących afiksów nieokreślonych

pozwala nam na wyjaśnienie zaobserwowanych form synkretyzmu przy założeniach teorii składni znanej jako nanoskładnia (Starke 2009, 2011). Najważniejsze elementy tej teorii to Superset Principle, Elsewhere Principle i frazowa leksykalizacja struktury składniowej. Zaobserwowane formy synkretyzmu powstają, gdy jedna forma fonologiczna realizuje więcej niż jedną strukturę gramatyczną (Superset Principle). Struktury, które mogą być wyrażone za pomocą jednej formy fonologicznej, muszą jednocześnie odpowiadać przyległym elementom w paradygmacie (Elsewhere Principle), co wyklucza powstanie synkretyzmu typu ABA. W rozprawie argumentuję, że forma synkretyzmu widoczna w paradygmacie twierdzących afiksów nieokreślonych w danym języku zależy od liczby form fonologicznych, które leksykalizują zaproponowaną hierarchię cech gramatycznych.

References

Abbott, Barbara. 2008. "The indefiniteness of definiteness", in: Thomas Gamerschlag, Doris Gerland, Rainer Osswald, Wiebke Petersen (eds.), *Frames and concept types.* Springer, 323-341.

Abney, Steven Paul. 1987. The English Noun Phrase in its Sentential Aspect. [Unpublished Ph.D. dissertation, MIT].

Aboh, Enoch Oladé. 2004. *The Morphosyntax of Complement-Head Sequences*. New York: Oxford University Press.

Aldridge, Maurice V. 1982. *English quantifiers: A study of quantifying expressions in linguistic science and modern English usage*. Amersham: Avebury.

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

Anchorena, José Dionisio. 1874. *Gramática Quechua ó del idioma del imperio de los incas.* Lima: Imprenta del Estado.

Bargery, George Percy and Westermann, Diedrich. 1934. *A Hausa-English dictionary and English-Hausa vocabulary*. London: Oxford University Press.

Baunaz, Lena and Eric Lander. 2018a. "Ontological categories", in: *The Unpublished Manuscript: a collection of Lingbuzz papers to celebrate Michal Starke's 50th birthday*. LingBuzz, 1-18. (https://lingbuzz.net/lingbuzz/003993) (date of access: 25 May 2022).

Baunaz, Lena and Eric Lander. 2018b. "Deconstructing categories syncretic with the nominal complementizer", *Glossa: A Journal of General Linguistics* 3, 1.

Baunaz, Lena and Eric Lander. 2018c. "Nanosyntax: the basics", in: Lena Baunaz, Karen De Clercq, Liliane Haegeman and Eric Lander (eds.), *Exploring Nanosyntax*. Oxford University Press, 13-84.

Baunaz, Lena, Karen De Clercq, Liliane Haegeman and Eric Lander (eds.). 2018. *Exploring*

nanosyntax. (Oxford Studies in Comparative Syntax.) Oxford University Press.

Belletti, Adriana (ed.). 2004. *Structures and beyond: The cartography of syntactic structures*, vol. 3. New York: Oxford University Press.

Bhat, Shankara. 1981. "Physical Identification in Kannada", *Studies in the linguistic sciences* 11, 2: 1-8.

Bliese, Loren. 1981. *A Generative Grammar of Afar*. (SIL International Publications in Linguistics 65.) Arlington: Summer Institute of Linguistics and the University of Texas.

Bobaljik, Jonathan. 2007. On comparative suppletion. (https://ling.auf.net/lingbuzz/000443) (date of access: 25 May 2022).

Bobaljik, Jonathan. 2012. *Universals in comparative morphology: suppletion, superlatives, and the structure of words*. Cambridge: MIT Press.

Bolinger, Dwight. 1977. "Any and some", in: Dwight Bolinger, *Meaning and Form* (English Language Series 11). London and New York: Longman, 21-36.

Bruening, Benjamin. 2007. "Wh-in-situ does not correlate with wh-indefinites or question particles", *Linguistic Inquiry* 38: 139-66.

Caha, Pavel. 2009. Nanosyntax of case. [Unpublished Ph.D. dissertation, University of Tromsø].

Caha, Pavel. 2017. "How (not) to derive a *ABA: The case of Blansitt's generalisation", *Glossa: a journal of general linguistics* 2, 1: 1–32.

Caha, Pavel. 2020. Nanosyntax: some key features. [Submitted as a competing theories chapter for *The handbook of distributed morphology*]. (https://ling.auf.net/lingbuzz/004437) (date of access: 25 May 2022).

Caha, Pavel. 2021. Minimalism and nanosyntax. [To appear in Kleanthes Grohmann and Evelina Leivada (eds.), *The Cambridge handbook of minimalism*. Cambridge University Press].

Caha Pavel, Taraldsen-Medova Lucie. 2020. "Syncretism in nanosyntax and three types of passive participles in Czech", *Naše řeč* 103: 113–130.

Chen, Zhuo. 2017. "The epistemic indefinite shenme in Mandarin", in: *Proceedings of the 34th West Coast Conference on Formal Linguistics*. Somerville: Cascadilla Proceedings, 115-121.

Cinque, Guglielmo. 1999. *Adverbs and inflectional heads*. Oxford: Oxford University Press.

Cinque, Guglielmo (ed.). 2002. *Functional structure in DP and IP*. (The Cartography of Syn-

tactic Structures, Vol. 1.) New York: Oxford University Press.

Cinque, Guglielmo. 2006. *Restructuring and functional heads*. (The Cartography of Syntactic Structures, Vol. 4.) New York: Oxford University Press.

Cinque, Guglielmo. 2008. "More on the indefinite character of the head of restrictive relatives", *Rivista di Grammatical Generativa* 33: 3-24.

Cinque, Guglielmo. 2010. *The syntax of adjectives*. Cambridge: MIT Press.

Cinque, Guglielmo. 2020. *The syntax of relative clauses*. (Cambridge Studies in Linguistics.) Cambridge University Press.

Cinque, Guglielmo and Luigi Rizzi. 2008. "The cartography of syntactic structures", *CISCL Working Papers on Language and Cognition* 2: 43-59.

Chomsky, Noam. 2001. "Derivation by phase", in: Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*. Cambridge: MIT Press, 1-52.

Corpus of Contemporary American English. (https://www.english-corpora.org/coca/) (date of access: 25 May 2022).

Croft, William. 1983. "Quantifier scope ambiguity and definiteness", *Berkeley Linguistic Society* 9, 25-36.

Cusihuamán, Antonio. 2001. *Gramática Quechua Cuzco-Collao*. Cuzco: Centro de estudios regionales andinos bartolomé de las casas.

De Bhaldraithe, Tomás (ed.). 1959. *English-Irish dictionary*. Dublin: Oifig an tSoláthair.

De Clercq, Karen. 2013. A unified syntax of negation. [Unpublished Ph.D. dissertation, University of Ghent].

De Clercq, Karen. 2020. *The morphosyntax of negative markers: a nanosyntactic account.* Berlin/Boston: Mouton De Gruyter.

De Clercq, Karen and Guido Vanden Wyngaerd. 2019. "On the idiomatic nature of unproductive morphology", *Linguistics in the Netherlands* 36, 1: 99-114.

Dekier. Jakub. 2021. "Morphosyntax of specific and non-specific indefinite markers", *Glossa: a journal of general linguistics* 6, 1: 70.

de Rijk, Rudolf. 2008. *Standard Basque: a progressive grammar*. Cambridge: MIT Press.

Dixon, Robert. M. W. 1977. *A Grammar of Yidiɲ.* Cambridge: Cambridge University Press.

Enç, Mürvet. 1991. "The semantics of specificity", *Linguistic Inquiry* 22, 1: 1-25.

Eremina, Olga. 2012. The Semantics of Russian indefinite pronouns: scope, domain widening, specificity, and proportionality and their interaction. [Unpublished Ph.D. dissertation,

Michigan State University].

Espinoza, Darío. 1997. *Tanteo puntun chaykuna valen.* Lima: CHIRAPAQ-Centro de Culturas Indias.

Fábregas, Antonio. 2007. "An Exhaustive lexicalisation account of directional complements", *Nordlyd* 34, 2: 165-199.

Faller, Martina. 2019. "A non-veridical indefinite in Cuzco Quechua", in: Kimberly Johnson and Alexander Göbel (eds.), *Proceedings of the Tenth Conference on the Semantics of Under-Represented Languages in the Americas: SULA 10*. University of Massachusetts Amherst, Graduate Linguistics Student Association, 17-32.

Faller, Martina. 2020. "The many functions of Cuzco Quechua =pas: implications for the semantic map of additivity", *Glossa: A Journal of General Linguistics* 5, 1: 34.

Frawley, William (ed.). 2003. *International encyclopedia of linguistics*. Oxford University Press.

Geist, Ljudmila. 2008. "Specificity as referential anchoring: evidence from Russian", in: Atle Grønn (ed.), *Proceedings of SuB12*. Oslo: ILOS 2008, 151-164.

Giannakidou, Anastasia. 1995. "Subjunctive, habituality and negative polarity", in Mandy Simons and Teresa Galloway (eds.), *Semantics and Linguistic Theory*(SALT) V. Ithaca: CLC Publications, Cornell University, 94-112.

Giannakidou, Anastasia. 2001. "The meaning of free choice", *Linguistics and Philosophy* 24: 659-735.

Giannakidou, Anastasia 2011a. "Nonveridicality and mood choice: subjunctive, polarity, and time", in: Renate Musan and Monika Rathert (eds.), *Tense across languages*. Berlin: Mouton de Gruyter.

Giannakidou, Anastasia. 2011b. "Positive polarity items and negative polarity items: variation, licensing, and compositionality", in: Claudia Maienborn, Klaus von Heusinger and Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*. Mouton de Gruyter, 1660-1712.

Giannakidou, Anastasia and Hedde Zeijlstra. 2017. "The landscape of negative dependencies: negative concord and n-words", in: Martin Everaert and Henk van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax, second edition*. John Wiley Sons, Inc, 1-38.

Gianollo, Chiara. 2018. *Indefinites between Latin and Romance*. (Oxford Studies in Diachronic and Historical Linguistics.) Oxford: Oxford University Press.

Giusti, Giuliana. 1997. "The categorical status of determiners", in: Liliane Haegeman (ed.), *The new comparative syntax*. New York: Longman Publishing Group, 95-123.

Grice, Paul. 1975. "Logic and conversation", in: P. Cole and J. J. Morgan (eds.), *Syntax and semantics 3: Speech acts*. New York: Academic Press, 41–58.

Hale, Kenneth. and Samuel J. Keyser. 1993. "On argument structure and the lexical expression of syntactic relations", in: Kenneth Hale and Samuel J. Keyser (eds.), *The view from building 20: A festschrift for Sylvain Bromberger*. Cambridge: MIT Press, 53-108.

Haegeman, Liliane. 1995. *The Syntax of Negation*. Cambridge: Cambridge University Press.

Haegeman, Liliane. 2006. "Argument fronting in English, Romance CLLD and left periphery.", in: Raffaella Zanuttini, Hector Campos, Elena Herburger, and Paul Portner (eds.), *Cross-linguistic research in syntax and semantics: negation, tense and clausal architecture*. Washington, D.C.: Georgetown University Press, 27-52.

Haegeman, Liliane 2012. *Adverbial clauses, main clause phenomena and the composition of the left periphery*. (The Cartography of Syntactic Structures, Vol. 8.) Oxford: Oxford University Press.

Haegeman, Liliane and Raffaella Zanuttini. 1991. "Negative heads and the neg criterion", *The Linguistic Review* 8: 233-51.

Haspelmath, Martin. 1993. *A Grammar of Lezgian*. (Mouton Grammar Library 9.) Berlin: Mouton de Gruyter.

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

Haspelmath, Martin. 2003. "The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison", in: Tomasello, Michael (ed.), *The new psychology of language*, vol. 2. Mahwah: Lawrence Erlbaum, 211-242.

Haspelmath, Martin and Josephine Caruana. 1996. "Indefinite pronouns in Maltese", *Rivista di Linguistica (Italian Journal of Linguistics)* 8, 1: 213-227.

Hawkins, John. 2015. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Routledge.

Hengeveld, Kees, Sabine Iatridou and Floris Roelofsen. 2021. "Quexistentials and focus", *Linguistic Inquiry*: 1–83. Massachusetts Institute of Technology.

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

Hewitt, George. 1996. *Georgian: A learner's grammar*. London: Routledge.

Hienonen, Heini. 2010. “The implicational semantic map for veps indefinite pronouns”, *Linguistica Uralica* 46, 4: 281-292.

Holton, David, Peter Mackridge and Irene Philippaki-Warburton. 2004. *Greek: An essential grammar of the modern language*. London: Routlege.

Hualde, José I. 1992. *Catalan*. London: Routledge.

Jabłońska Patrycja. 2007. Radical decomposition and argument structure. [Unpublished Ph.D. dissertation, CASTL/University of Tromsø].

Jackendoff, Ray. 1983. *Semantics and cognition*. Cambridge: The MIT Press.

Jenkins, E. S. 1961. *French grammar*. London: The English Universities Press.

Jónsdóttir, M. 1991. “On indefinite pronouns in Icelandic”, in: Halldór Á. SigurDsson (ed.), *Papers from the Twelfth Scandinavian Conference of Linguistics*. Reykjavík: Institute of Linguistics, University of Iceland, 153-70.

Kaiser, Stefan, Yasuko Ichikawa, Noriko Kobayashi and Hilofumi Yamamoto. 2001. *Japanese: A comprehensive grammar*. London: Routledge.

Karttunen, Lauri. 1976. “Discourse referents”, in: J. D. McCawley (ed.), *Syntax and semantics* 7. Academic Press, 363-386.

Kayne, Richard S. 1983. “Connectedness”, *Linguistic Inquiry* 14: 223-249.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge: MIT Press.

Kayne, Richard. 2005. *Movement and silence*. (Oxford studies in comparative syntax.) Oxford University Press.

Klima, Edward E. 1964. “Negation in English”, in: Jerry A. Fodor and Jerrold J. Katz (eds.), *The Structure of language*. Englewood Cliffs: Prentice-Hall Inc, 246-323.

Korpela, Jukka K. 2015. *Handbook of Finnish*. Suomen E-painos Oy.

Koul, Omkar. 2008. *Modern Hindi Grammar*. Springfield: Dunwoody Press.

Kozhanov, Kirill. 2015. “Lithuanian indefinite pronouns in contact”, in Peter Arkadiev, Axel Holvoet and Björn Wiemer (eds.), *Contemporary approaches to Baltic linguistics*. Mouton de Gruyter, 465-490.

Lakoff, Robin. 1969. “Some reasons why there can’t be any some-any rule”, *Language* 45: 608-15.

Lander, Eric and Liliane Haegeman. 2016. “The nanosyntax of spatial deixis”, *Studia Linguistica* 72, 2: 362-427.

Lawler, John M. 1971. “any Questions?”, *Chicago Linguistic Society* 7: 163-80.

Lavric, Eva. 1990. *Missverstehen verstehen: Opake Kontexte und Ambiguitäten bei indefiniten und definiten Nominalphrasen* (Grazer Linguistische Monographien 7). Graz: Institut für Sprachwissenschaft der Universität Graz.

Lewis, Charlton T. 1891. *An elementary Latin dictionary.* New York: Harper Brothers.

Li, Yen-Hui Audrey. 1992. "Indefinite wh in Mandarin Chinese", *Journal of East Asian Linguistics* 1, 2: 125-55.

Lin, Jo-Wang. 1998. "On existential polarity WH-phrases in Chinese", *Journal of East Asian Linguistics* 7, 3: 219-255.

Lukeš, Dominik and Jitka Kauerová. 2012. *Yellow pages of the Czech language: Grammar and usage for learners.* Lulu.com.

Markan evangelii. 1992. Saarijärvi: Biblijan Kändamižen Institut.

Ma Newman, Roxana. 1990. *An English-Hausa Dictionary*. New Haven: Yale University Press.

Martin, Samuel E. and Young-Sook Lee. 1986. *Beginning Korean*. Rutland: Tuttle.

Matsumoto, Yoshiko. 1997. *Noun-modifying constructions in Japanese: A frame-semantic approach*. John Benjamins Publishing.

McCawley, James D. 1977. "Lexicographic notes on English quantifiers", *Chicago Linguistic Society* 13: 372-83.

McCulloch, Gretchen. 2012. Mi'gmaq Indefinite Pronouns. [Paper presented at Oklahoma Workshop on Native American Languages, April 2012].

Mpiranya, Fidèle. 2015. *Swahili Grammar and Workbook*. Abignon: Routledge.

Nguyen, Tam. 2013. A grammar of Bih. [Unpublished PhD dissertation, University of Oregon]. (hdl.handle.net/1794/12996) (date of access: 25 May 2022).

Noonan, Michael. 1992. *A Grammar of Lango*. (Mouton Grammar Library 7.) Berlin: Mouton de Gruyter.

Ó Dónaill, Niall (ed.) 1977. *Foclóir gaeilga-béarla*. Dublin: Oifig an tSoláthair.

Orlandini, Anna. 1981 "Semantica e pragmatica dei pronomi indefiniti latini", *Lingua e Stile* 16, 2: 215-34.

Pantcheva, Marina. 2008. "The place of PLACE in Persian", in: Anna Asbury, Jakub Dotlačil, Berit Gehrke, and Rick Nouwen (eds.), *Syntax and semantics of spatial P*. (Linguistik Aktuell 120.) John Benjamins, 305–330.

Pantcheva, Marina. 2011. Decomposing path: The nanosyntax of directional expressions. [Un-

published Ph.D. dissertation, Tromsø: University of Tromsø].

Parker, Enid. 1974. *Afar/English dictionary, Part III: modifiers associated with Class III verbs*. Addis Ababa: Red Sea Mission Team.

Parker, Gary. 1976. *Gramática Quechua, Ancash-Huailas*. Lima: Ministerio de education.

Patrick, Jon and Ilari Zubiri Ibarrondo. 2001. *A Student Grammar of Euskara*. LINCOM publishers.

Pollock, Jean-Yves. 1989. "Verb movement, universal grammar, and the structure of IP.", *Linguistic Inquiry* 20, 3: 365-424.

Postma, Gertjan. 1994. "The indefinite reading of WH", in: Reineke Bok-Bennema and Crit Cremers (eds.), *Linguistics in the Netherlands* 1994. Amsterdam: Benjamins, 187-198.

Proudfoot, Anna and Francesco Cardo. 1997. *Modern Italian grammar: A practical guide*. Routledge.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Harlow: Longman.

Ramchand, Gillian. 2008. *Verb meaning and the lexicon: A first-phase syntax*. Cambridge: Cambridge University Press.

Richtarcikova, Veronika. 2013. "Epistemic indefinites in Slovak: Alternatives and exhaustification?", *10th European Conference on Formal Description of Slavic Languages (FDSL-10)*. Germany: University of Leipzig.

Rizzi, Luigi. 1997. "The fine structure of the left periphery", in: Liliane Haegeman (ed.), *Elements of Grammar*. Dordrecht: Springer, 281-337.

Rizzi, Luigi. 2001. "On the position int(errogative) in the left periphery of the clause", in: Guglielmo Cinque and Giampaolo Salvi (eds.), *Current Studies in Italian Syntax: Essays Offered to Lorenzo Renzi*. Amsterdam: Elsevier, 287-296.

Rizzi, Luigi. (ed.). 2004. *The Structure of CP and IP*. (The Cartography of Syntactic Structures, Vol. 2.) New York: Oxford University Press.

Roberts, John. 2003. Persian Grammar Sketch. (http://tulquest.huma-num.fr/sites/default/files/questionnaires/89/Roberts$_{PersianGrammarSketch.pdf}$)($dateofaccess$: 25May2022).

Rounds, Carol. 2001. *Hungarian: An essential grammar*. London: Routlege.

Sharahsenidze, Nino. 2018. "Teaching Georgian as a second language: indefinite pronouns and modality", *International Journal of Multilingual Education* 11: 126-128.

Shimelman, Aviva. 2017. *A grammar of Yauyos Quechua*. (Studies in Diversity Linguistics 9.)

Language Science Press.

Shlonsky, Ur. 2010. “The Cartographic Enterprise in Syntax”, *Language and Linguistics Compass* 4, 6: 417-429.

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”, *Nordlyd* 36, 1: 1-6.

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* 39: 245-256.

Starke, Michal. 2018. “Complex left branches, spellout and prefixes”, in: Lena Baunaz, Karen De Clercq, Liliane Haegeman and Eric Lander (eds.), *Exploring Nanosyntax*. Oxford University Press, 349-363.

Takamine, Kaori. 2007. “The Axial Part Phrase in Japanese”. *Nordlyd* 33: 78–97.

Taraldsen, Tarald. 2010. “The nanosyntax of Nguni noun class prefixes and concords”, *Lingua* 120: 1522-1548.

Taraldsen, Knut T., Lucie Taraldsen Medová and David Langa. 2018. “Class prefixes as Specifiers in Southern Bantu”, *Natural Language Linguistic Theory* 36, 4: 1339-1394.

Taraldsen-Medova, Lucie and Bartosz Wiland. 2018. “Functional sequence zones and Slavic L>T>N participles”, in: Lena Baunaz, Karen De Clercq, Liliane Haegeman and Eric Lander (eds.), *Exploring Nanosyntax*. New York: Oxford University Press, 305–328.

Taraldsen-Medova, Lucie and Bartosz Wiland. 2019. “Semelfactives are bigger than degree achievements”, *Natural Language and Linguistic Theory* 37, 4: 1463–1513.

Tolskaya, Inna. 2018. “Nanosyntax of Russian verbal prefixes”, in: Lena Baunaz, Karen De Clercq, Liliane Haegeman and Eric Lander (eds.), *Exploring Nanosyntax*. New York: Oxford University Press, 205–236.

Türk, Utku and Pavel Caha. 2021. “Nanosyntactic analysis of Turkish case system”, *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic* 6. (https://ling.auf.net/lingbuzz/005946) (date of access: 25 May 2022).

Warfel, Sam L. 1972. “Some, reference, and description”, in: J. H. Battle and John Schweitzer (eds.), *Mid-America Linguistics Conference Papers*. Stillwater: Oklahoma State Uni-

versity, 41-9.

Wągiel Marcin, Pavel Caha. 2021. “Complex simplex numerals”, *Acta Linguistica Academica* 68, 4: 470–515.

Weber, David, John. 1989. *A Grammar of Huallaga (Huanuco) Quechua*. University of California Press.

White, Leila. 2006. *A grammar book of Finnish*. Helsinki: Oy Finn Lectura Ab.

Wiland, Bartosz. 2012. “Prefix stacking, syncretism, and the syntactic hierarchy”, in: Mojmir Docekal and Marketa Zikova (eds.), *Slavic Languages in Formal Grammar*. Frankfurt am Main: Peter Lang, 307–324.

Wiland, Bartosz. 2018. “A note on lexicalizing ‘what’ and ‘who’ in Russian and in Polish”, *Poznan Studies in Contemporary Linguistics* 54, 4: 573–604.

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

Windisch, Ernst. 1882. *Compendium of Irish grammar*. (Translated by James P. M’Swiney.) Dublin: M. H. Gill & son.

Vanden Wyngaerd, Guido. 2017. “The feature structure of pronouns: A probe into multidimensional paradigms”, in: Lena Baunaz, Karen De Clercq, Liliane Haegeman and Eric Lander (eds.), *Exploring Nanosyntax*. Oxford University Press, 277-304.

Vogt, Hans. 1971. *Grammaire de la langue géorgienne*. Oslo: Universitetsforlaget.

Secondary sources

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

Dončeva, Liljana. 1970. "Ruskite neopredelitelni mestoimenija i narečija ot tipa -to, -nibud'(libo), ... ni bylo i texnite ekvivalenti v bâlgarski ezik", *Bâlgarski ezik* 1970, 5: 442-51.

Kodima. 1999. 11-12. Petroskoi.

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

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

Paducheva, Elena V. 1985. *Proposition and its relation to reality: referential aspects of pronouns semantics*. Moscow: Science.

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

Saltarelli, Mario. 1988. *Basque*. (Routledge Descriptive Grammars.) London: Routledge.

Setälä, Emil, Nestor and J. H. Kala. 1951. *Näytteitä äänis-ja keskivepsän murteista*. Helsinki: Suomalais-ugrilainen Seura.

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

Appendix

The appendix contains the data collected for the purpose of the proposed analysis and their sources. For each language, I provide examples illustrating the indefinite pronoun forms used in the non-specific, specific unknown and specific known functions. Some of the native speakers who I interviewed did not want their names to be revealed, which is why I list them as anonymous sources.

Full syncretism

(258) English (Haspelmath 1997)

a. *Let's go **somewhere*** (non-specific).

b. *There is **someone** behind the door* (specific unknown).

c. *I have **something** for you* (specific known).

For more information on English indefinites see Aldridge (1982), Bolinger (1977), Lakoff (1969), Lawler (1971), McCawley (1977), Quirk et al. (1985).

(259) Polish (data according to my own native knowledge of the language)

a. *Przynieś mi **co-ś** do jedzenia.*
bring.IMPV I.DAT what.ACC-INDEF to eating.GEN
'Bring me **something** (non-specific) to eat.'

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** called. Guess who.'

(260) Japanese (Haspelmath 1997: 312, Kaiser et al. 2001, my own knowledge of the language and an anonymous native speaker)

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

b. ***Dare-ka*** *kara 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*** *kara 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.'

(261) Korean (Haspelmath 1997: 314-315, Martin and Lee 1986)

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.'

According to Haspelmath (1997: 314-315), the indefinite morpheme *-nka* can be omitted in colloquial speech.

(262) French (Haspelmath 1997:315, an anonymous native speaker)

a. ***Quelqu'un*** *m'attend.* *C'est lui que j'épouserai.*
someone is.waiting.for.me it.is him that I.will.marry
'**Someone** is waiting for me. It's him that I will marry.'

b. ***Quelqu'un*** *est à la porte. Je ne sais pas qui.*
someone is at the door I neg know neg who
'**Somebody** is at the door. I don't know who.'

c. *Elle veut épouser* ***quelqu'un*** *aux cheveux noirs.*
she wants marry someone at.the hair black
'She wants to marry **someone** (non-specific) with black hair.'

(263) Dutch (Haspelmath 1997: 247, Hengeveld et al. 2020, an anonymous native speaker)

a. *Er heeft vanmorgen* ***iemand*** *opgebeld.* (*Raad eens wie./ Ik weet niet meer wie het was.*)
there has this.morning someone phoned guess once who I know neg more who it was
'**Someone** phoned this morning. (Guess who./I don't remember who it was.)'

b. *Koop* ***iets*** *voor haar verjaardag.*
buy something for her birthday
'Buy **something** (non-specific) for her birthday.'

c. *Miranda heeft* ***wat*** *gegeten.*
Miranda has what eaten
'Miranda has eaten **something**.'

d. *Ik hab* ***wat*** *gegeten.*
I have what eaten
'I have eaten **something**.'

e. *Hij zit steeds over* ***iets/wat*** *na te denken.*
he sits constantly about something/what after to think
'He is constantly thinking about **something**.'

Haspelmath (1997:246) suggests that the *-dan ook* series can appear in the non-specific function as well as in the free choice function. However, a native speaker I interviewed treats *-dan ook* forms only as free choice indefinites.

(264) Sweedish (Haspelmath 1997: 250, an anonymous native speaker)

a. *Jag har ännu* ***något*** *att tala med honom om.*
I have still something to talk with him about
'I still have **something** (specific known) to talk to him about.'

b. *Det finns* ***något*** *i lådan*
there is something in the.box
'There is **something** (specific unknown) in the box.'

c. *Hämta mig* ***något*** *att äta*
fetch me something to eat
'Bring me **something** (non-specific) to eat.'

(265) Finnish (Hasplemath 1997: 293, Korpela 2015, an anonymous native speaker)

a. ***Joku*** *soitt-i. Arvaa kuka* (*se oli*)?
someone call-PAST(3SG) guess.IMPV who it was
'**Someone** (specific known) called. Guess who it was.'

b. ***Joku*** *soitt-i, mutta en saa-nut nime-stä*
someone call-PAST(3SG) but NEG.1SG get-PAST.PTCP name-ELAT
selvää.
clear.PRTV
'**Someone** (specific unknown) called, but I didn't understand the name.'

c. *Hän haluaa mennä naimisiin* ***jonkun*** *kanssa*, *jolla on tumma tukka.*
she wants go marriage someone with who-on is dark hair
'She wants to marry **someone** (non-specific) with black hair.'

According to Haspelmath (1997: 293-294), the specific known function in Finnish does not have a corresponding indefinite pronoun series. A native speaker I interviewed considers the basic *joku*/*jokin* series grammatical in the specific known function.

(266) Greek (Haspelmath 1997: 266, Holton et al. 2004: 96-98, Elissavet Antoniadou and two other anonymous sources.)

a. ***Ká-pjos*** *tilefónise. Mándepse pjos!*
INDEF-who phoned guess.IMPV who
'**Someone** (specific known) called. Guess who!'

b. ***Ká-pjos*** *tilefónise. Dhen kséro pjos.*
INDEF-who phoned not know.1P.SG who
'**Someone** called (specific unknown). I don't know who.'

c. *Fére* ***ká-ti*** *na fáme!*
bring.IMPV INDEF-what SBJV eat.2P.PL
'Bring **something** (non-specific) to eat!'

According to Haspelmath (1997: 256-266), the *ka-* series cannot be used in the specific known function. Native speakers I interviewed judged *ka-* forms grammatical in the specific known function.

(267) Basque (Haspelmath 1997: 315-316, Patrick and Ibarrondo 2001, Rijk 2008, Saltarelli 1988)

a. *Kanpoa-n zengoze-n-en-ean* ***baten batek*** *dei egin zizun. Esan nor.*
away-LOC be-2SG-REL-LOC one.GEN one.ERG call make he.you.has say who
'**Someone** (specific known) called while you were away. Guess who it was.'

b. *Hiretzat ere aukeratuta neukan* ***norbait***.
for.you too chosen I.had who-INDEF
'For you too I had chosen **someone** (specific known?).'

c. ***Nor-bait-i*** *utzi nion argazki-makina eta ez naiz oroi-tzen*
who-INDEF-DAT lend I.it.to.him photo-machine and NEG I.it remember-HAB
nor-i.
who-DAT
'I lent the camera to **someone** (specific unknown) and I do not remember to whom.'

d. ***Nora-bait*** *joa-te-ko gogoa dut, baina ez dakit nora.*
where-INDEF go-HAB-REL thought I.have.it but NEG I.know.it where
'I have an urge to go **somewhere** (non-specific) but I do not know where.'

According to Haspelmath (1997: 315), the specific known function in Basque cannot be expressed with indefinite pronouns. An expression with the determiner *bat* 'one' combined with a noun may be used to describe specific known referents. However, other sources such as Patrick and Ibarrondo (2001) and Rijk (2008) do not mention the ungrammaticality of *-bait* indefinites in the specific known function.

(268) Icelandic (Haspelmath 1997: 252, Jónsdóttir 1991)

a. *Það hringdi* ***ein-hver*** *í mig í morgun* (*en ég segi þér ekki hver*).
it rang INDEF-who in me in morning and I say you not who
'**Someone** (specific known) called me up this morning (but I won't tell you who).'

b. *Strákur-inn sá* ***eitt-hvað*** *á bak við tréð* (*en ég veit ekki hvað*).
boy-the saw INDEF-what on back near tree and I know not what
'The boy saw **something** (specific unknown) behind the tree (but I don't know what).'

c. *Kaupta* ***eitt-hvað*** *fyrir mig!*
buy INDEF-what for me
'Buy **something** (non-specific) for me!'

(269) Portuguese (Haspelmath 1997: 256)

a. *O João telefonou e disse* ***alguma*** *coisa. Adivinhe o que!*
ART João phoned and said some thing guess what
'João called and said **something** (specific known) – guess what!'

b. *Senti que* ***alguém*** *me observava.*
I.felt that someone me observed
'I felt that **someone** (specific unknown) was observing me.'

c. ***Alguém*** *pode passar aí e apanhar a caixa.*
someone can pass here and take the box
'**Someone** (non-specific) can come along here and take the box.'

(270) Catalan (Haspelmath 1997: 258, Hualde 1992)

a. *Alguna vegada he vist* ***algú*** ***en algun lloc***.
some time I.have seen someone in some place
'I once saw **somebody somehwere** (specific unknown/known).'

b. *Algun* ***dia*** *ho entendràs.*
some day it you'll.understand
'**Some day** you'll understand (non-specific).'

According to Haspelmath (1997: 258), the *algú* series may be used in the non-specific, specific unknown and specific known indefinite functions. The source does not provide examples for the non-specific function with a pronominal form.

(271) Italian (Haspelmath 1997: 262-263, Proudfoot and Cardo 1997, native speaker: Maria Cortiula)

a. ***Qualcuno*** *è venuto – indovina chi!*
some-one is came guess who
'**Someone** (specific known) has come - guess who!'

b. *C'è* ***qualcuno*** *alla porta.*
there.is some-one at.the door
'There is **somebody** (specific unknown) at the door.'

c. *Compra* ***qualcosa*** *per me.*
buy some-thing for me
'Buy **something** (non-specific) for me.'

(272) Lezgian (Haspelmath 1993, Haspelmath 1997: 296)

a. *Za wa-z* ***sa wuč jat'ani*** *luhu-da.*
I.ERG you-DAT one what-INDEF say-FUT
'I'll tell you **something** (specific known).'

b. *Gilač ar-ar* ***wuč iz jat'ani*** *xükwe-zma-č.*
now letter-PL why-INDEF come-CONT.IMPF-NEG
'Now the letters don't come anymore **for some reason** (specific unknown).'

c. *Sa* ***ni-z jat'ani*** *ewer ce!*
one who-DAT-INDEF call give.IMPV
'Call **someone** (non-specific)!'

Lezgian indefinites may be preceded by *sa* 'one'.

(273) Romanian (Haspelmath 1997: 265)

a. *Am cumpărat* ***ceva****, ghici ce!*
I.have bought what-INDEF guess what
'I have bought **something** (specific known), guess what!'

b. ***Cineva*** *te-a căutat la telefon, dar nu i-am înțeles numele.*
who-INDEF you-has asked on phone but not him-have understood name
'**Somebody** (specific unknown) asked for you on the phone, but I didn't understand the name.'

c. *Aș vrea* ***ceva*** *mâncare și băutură.*
INTJ want what-INDEF to.eat and drink
'I want **something** (non-specific) to eat and to drink.'

(274) Bulgarian (Haspelmath 1997: 268, Dončeva 1970: 445 as cited in Haspelmath 1997: 268)

a. *Govorix včera s* ***nja-koga*** *v avtobusa.*
I.talked yesterday with INDEF-whom in the.bus
'Yesterday I talked to **someone** (specific unknown) on the bus.'

b. ***Nja-koj*** *se obadi, otgatni koj.*
INDEF-who REFL called guess.IMPV who
'**Somebody** has called, guess who.'

c. *Kogato edin čovek iska da vârši* ***ne-što*** *- daj mu rabota.*
when one person wants SBJV he.do INDEF-what give.IMPV him work
'When a person wants to do **something** (non-specific) – give him or her work.'

(275) Serbo-Croatian (Haspelmath 1997: 270)[78]

a. *Mira voli* ***ne-koga***.
Mira loves INDEF-whom
'Mira loves **someone** (specific unknown/known).'

b. *Daj mi* ***ne-što!***
give me INDEF-what
'Give me **something** (non-specific)!'

(276) Latvian (Haspelmath 1997: 277)

a. *Viņš* ***kaut-kur*** *aizgājis*.
he INDEF-where left
'He went **somewhere** (specific unknown/known).'

b. *Nopērc* ***kaut-ko***.
buy INDEF-what
'Buy **something** (non-specific).'

(277) Hindi (Haspelmath 1997: 284, Koul 2008: 79)

a. ***Kisii-ne*** *fon kiy-aa thaa*, - *par mãĩ tumh nahĩ bataaũgii*, *kis-ne*.
someone-ERG phone do-PFV was but I you NEG I.will.tell who-ERG
'**Someone** has phoned, but I won't tell you who.'

b. ***Kisii-ne*** *fon kiy-aa thaa*, - *par mujhe nahĩ maaluum*, *kis-ne*.
someone-ERG phone do-PFV was but I.DAT NEG known who-ERG
'**Someone** has phoned, but I don't know who'

c. ***Kisii-ko*** *fon kar-o!*
someone-DAT phone do-IMPV
'Phone **someone** (non-specific)!'

[78] Haspelmath (1997: 270) uses the term Serbo-Croatian to describe this set of data. I understand that it may be considered inaccurate.

(278) Turkish (Haspelmath 1997: 286-287)

a. *San-a* ***biri(si)*** *telefon et-ti.* *Bil*, *bakalım*, *kim.*
you-DAT one phone do-PAST know we'll.see who
'**Someone** phoned you. Guess who.'

b. *San-a* ***biri(si)*** *telefon et-ti.* *Kim ol-duğ-u-nu*
you-DAT someone phone do-PAST who be-REL-3SG-ACC
bil-m-iyor-um.
know-NEG-IMPF-1SG
'**Someone** phoned you. I don't know who.'

c. ***Bir*** *yer-e* *telefon et!*
one place-DAT phone do(IMPV)
'Call **somewhere** (non-specific)!'

(279) Kazakh (Haspelmath 1997: 288)

a. ***Birew*** *telefon soq-ty - byraq kim ekenin bil-me-j-min.*
someone phone hit-PAST but who [?] know-NEG-*pres*-1SG
'**Someone** called - but I don't know who.'

b. ***Birew*** *telefon soq-ty - kim ekenin tapšy.*
someone phone hit-PAST who [?] guess(IMPV)
'**Someone** called - guess who.'

c. ***Bir*** *žer-de* *kezde-ser-miz.*
one place-INESS meet-FUT-1PL
'Let's meet **somewhere** (non-specific).'

(280) Hungarian (Haspelmath 1997: 291, Rounds 2001, native speaker: Daniel Pap)

a. ***Vala-hol*** *már* *tolál-koz-t-unk.*
INDEF-where already meet-REFL-PAST-1PL
'We have met **somewhere** (specific unknown) before.'

b. *Mester*, *lát-ánk* ***vala-ki-t***, *aki a te nev-ed-del*
master see-1PL INDEF-who-ACC who ART you name-2SG-INSTR
ördög-ök-et üz.
devil-PL-ACC drive

‘Master, we saw **someone** (specific known) casting out devils in thy name.’

c. *Majd csak meglesz-ünk* ***vala-hogy***.
later only manage-1PL INDEF-how
‘We’ll manage **somehow** (non-specific).’

(281) Czech (Lukeš and Kauerová 2012: 34, native speakers: Michaela Faltýnková and Šárka Melicharová)

a. ***Něco*** *pro tebe mám. Hádej co?*
INDEF-what for you have.1P.SG gues-IMPV what
‘I have **something** for you. Guess what?’

b. ***Někdo/kdosi*** *volal. Nevím kdo.*
INDEF-who/who-INDEF called NEG.know who
‘**Someone** called. I don’t know who.’

c. *Přines mi* ***něco*** *k jídlu.*
bring-IMPV me INDEF-what for eating
‘Bring me **something** (non-specific) to eat’

Apart from the general *ně-* series, Czech speakers also use the *-si* series, which may only appear in the specific unknown function.

(282) Slovak (Richtarcikova 2013)

a. ***Niekoho/Vol’akoho/Kohosi*** *mi pripomína.*
someone.ACC me reminds
‘He/She reminds me of **somebody** (specific unknown).’

b. ***Niekoho*** *mi pripomína. Vieš koho? Môjho bývalého.*
someone.ACC me reminds know who my ex
‘He reminds me of **somebody** (specific known). Do you know who? My ex.’

c. *Pod, dušicka, zobni si* ***niečo/vol’ačo/čosi*** *nech neodpadneš.*
come soul peck PART something so NOT.faint
‘Come, darling, eat **something** (non-specific) so that you don’t faint.’

Apart from the general *nie-* series, Slovak also uses two specific unknown/non-specific series,

namely the *-si* and *vol'a-* series.

(283) Classical Greek (New Testament: Luke 8: 46, Acts 3: 5, Mark 9: 38)

a. *Éphē autōi ho Iōánnēs, Didáskale, eídomén* ***tina*** *en tōi onómatí sou*
said himself the John, teacher saw.1PL someone in the name your
ekbállonta daimónia.
cast.out demons
'Teacher, we saw **someone** (specific known) casting out demons in your name.'

b. *Ho dè Iēsoûs eîpen, Ēpsató moú* ***tis****, eg gàr égnōn dýnamin*
the but Jesus said, touched me someone I for know power
exelēlythyîan ap' emoû.
went.out from me
'But Jesus said, **"Someone** (specific unknown) touched me; I know that power has gone out from me."'

c. *Ho dè epeîchen autoîs prosdokōn* ***ti*** *par' autōn labeîn.*
the but gave.attention he expect something from them obtain
'So the man gave them his attention, expecting to get **something** (non-specific) from them.'

(284) Maltese (Haspelmath 1997: 299, Haspelmath and Caruana 1996)

a. ***Xi add*** *qed jadem fil-ġnien.*
someone PROG he.work,IMPF in.the-garden
'**Somebody** (specific unknown/known) is working in the garden.'

b. ***Xi*** *darba is-sena d-diela mmur l-Afrika.*
INDEF time the-year the-entering I.go.IMPF the-Africa
'I'll go to Africa **sometime** (non-specific) next year.'

(285) Hebrew (Haspelmath 1997: 297)

a. *Rai-ti* ***mi-šehu****.*
saw-1SG who-INDEF
'I saw **somebody** (specific unknown/known).'

b. *Ti-re* ***ma-šehu***.
2SG-will.see what-INDEF
'You will see **something** (non-specific).'

(286) Hausa (Haspelmath 1997: 300, Bargery and Westermann 1934)

a. *Àlbishirin-kà* ***wani*** *yaa* *kiraa kà*.
good.news.of-you someone 3SG.PAST call you
'**Someone** has called you, guess who?'

b. ***Wani*** *yaa* *kiraa kà*, *amma ban* *san* *koo wàa ba*.
someone 3SG.PAST call you but NEG.1SG know Q who NEG
'**Someone** has called you, but I don't know who.'

c. ***Wata*** *raanaa sáà* *daawoo*.
some day 3PL.FUT return
'**Some day** (non-specific) they'll return.'

(287) Colombian Spanish (native speaker: Daniela Osorio Castro)

a. *Traje* *a* ***alguien***. *Adivinia* *a* *quién*.
bring.1SG.PAST PREP someone guess.IMPV PREP who
'I have brought **someone** (specific known). Guess who.'

b. *Hay* ***alguien*** *detrás de* *la puerta*. *No* *sé* *quién es*.
there.is someone behind PREP the door NEG know.1SG who is
'There is **someone** (specific-unknown) behind the door. I don't know who.'

c. *Llama* *a* ***alguien***! *No* *importa* *a* *quién*.
call.IMPV PREP someone NEG matter.3SG PREP who
'Call **somebody** (non-specific)! It doesn't matter who.'

(288) German (Haspelmath 1997: 245, native speaker: Philip Wengel)

a. *Ich habe* ***etwas*** *verloren*. *Rate* *mal*, *was?*
I have something lost guess PART what
'I lost **something** (specific known). Guess what?'

b. *Ich habe **etwas/irgend-etwas** verloren, aber ich weiss nicht, was.*
I have something lost but I know not what
'I lost **something** (specific unknown), but I don't know what.'

c. *Sie möchte **jemanden/irgend-jemanden** mit schwarzen Haaren heiraten.*
She would.like someone with black hair marry
'She wants to marry **someone** (specific unknown/non-specific) with black hair.'

d. *Auf was für einen Film hast du Lust? Hmmm keine Ahnung...*
on what for a film have you desire hmm no idea
***irgend-was** lustiges.*
anything/something funny
'What kind of movie do you feel like watching? Hmmm ... no idea... anything funny/**something** (non-specific) funny.'

e. *Auf was für einen Film hast du Lust? Hmmm keine Ahnung... auf jeden*
on what for a film have you desire hmm no idea on every
*Fall **was/etwas** lustiges.*
case what funny
'What kind of movie do you feel like watching? Hmmm ... no idea... definitely **something** (non-specific) funny.'

f. *Kannst du mir **was** mitbringen?*
can you me what bring
'Can you bring me **something**?'

The *irgend-* series may appear only in the non-specific and specific unknown functions. When stressed *irgend-* indefinites receive a free choice interpretation. In colloquial speech, interrogative pronouns can be used as indefinites.

(289) Persian (Haspelmath 1997: 283, Roberts 2003)

a. Diruz *kas-i*-râ dar otobus did-am.
yesterday person-INDEF-ACC on bus see.PAST-1SG
'Yesterday I saw *someone* (specific known/unknown) on the bus.'

b. *Yek vaqt-i* be xâne-ye mâ bey-âyid.
one time-INDEF to house-LK we SBJV-come.IMPV
'Come to our house *sometime* (non-specific).'

Indefinite forms in Persian are generic nouns with the indefinite marker/article *-i*. It is not clear how grammaticalized these forms are, and if they may be considered indefinite pornouns.

(290) Mi'gmaq (McCulloch 2012)

a. *Natu-wen pegising'p*.
INDEF-who arrived
'Someone arrived.' (specific known)

b. *Natu-wen nutaqap*.
INDEF-who I.heard
'I heard someone.' (specific unknown)

c. *Na-tami amujpa-liedi*.
INDEF-where have.to-you.go
'You'll have to go somewhere (else).' (non-specific)

(291) Slovenian (native speaker: Gregor Kladnik)

a. *Imam* ***ne-kaj*** *zate*. *Ugani, kaj*.
have.1SG INDEF-what for.you guess what
'I have **something** for you. Guess what.'

b. *Mary se želi poročiti z* ***ne-kom*** *iz ZDA. Ne vem, kdo to je*.
Mary self wants marry with INDEF-who from US not know.1SG who that is
'Mary wants to marry **someone** from the US. I don't know who.'

c. *Prinesi mi* ***ne-kaj*** *za jesti. Vseeno mi je, kaj*.
bring me INDEF-what for eat anyway me is what
'Bring me **something** to eat. It doesn't matter what.'

No syncretism

(292) Lithuanian (Kozhanov 2015, Haspelmath 1997: 276, Pilka 1984 as cited in Haspelmath 1997)

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.'

(293) Russian (Eremina 2012, native speakers: Maria Molina and Maxim Orechnikov)

a. *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.'

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

c. *Ja tebe* ***koe-što*** *prines*. *Smotri*, *kakaja dynia*.
I you-DAT INDEF-what brought. Look which melon.
'I brought you **something** (specific known). Look at this melon.'

The *-nibud* marker may be replaced with *-libo* in formal contexts.

ABB syncretism

(294) Nanay (Haspelmath 1997: 67-68, Onenko 1986 as cited in Haspelmath 1997: 67-68)

a. ***Uj-nuu*** *žook-či* *laŋ* *ži-či-ni.*
who-INDEF house-DIR near go-PAST-3SG
'**Someone** (specific unknown/known) 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 unknown/known).'

c. ***Xaj-daa*** *osii-daa* o*si-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.'

(295) Georgian (Haspelmath 1997: 304, Hewitt 1996)

a. *Es* *c'igni* ***sad-γac*** *v-išove.*
this book where-INDEF 1SG-found
'I found this book **somewhere** (I could say where).'

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

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

(296) Ossetic (Haspelmath 1997: 281, Kulaev 1958 as cited in Haspelmath 1997: 281)

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.'

(297) Yakut (Haspelmath 1997: 290, Afans'ev and Xaritonov 1968 and Ubrjatova 1982 as cited in Haspelmath 1997: 290)

a. ***Kim-ere*** *eji-exe kiir-e syld'y-byt-a*.
who-INDEF you-DAT enter-COV 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).'

AAB syncretism

(298) Latin (New Testament: Luke 8: 46, Acts 3: 5, Mark 9: 38, Gianollo 2013, Haspelmath 1997: 254, Orlandini 1981)

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 ab eis*.
but that gave.heed in them hoping self INDEF-what accept.FUT from them
'And he gave heed unto them, expecting to receive **something** (non-specific) of them.'

c. *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.'

Incomplete paradigms

(299) Mandarin Chinese (Haspelmath 1997: 307-308, Li 1992, Lin 1998, native speaker: Wesley Wu):

a. *Chī, diǎn* ***shénme*** *zài zǒu ba!*
eat a.bit what then go PART
'Please eat a little **something** (non-specific) before you leave.'

b. *Yǒu* ***rén*** *dǎ diànhuà le. Wo bù zhīdao shì shéi.*
exist man hit phone PERF I not know is who
'**A person** (specific unknown) called. I don't know who.'

c. *Yǒu* ***rén*** *dǎ diànhuà le. Cāi yī cāi shì shéi.*
exist man hit phone PERF guess one guess is who
'**A person** (specific known) called. Guess who.'

(300) Kannada (Haspelmath 1997: 305-306, Bhat 1981)

a. ***Yaar-oo*** *bandaru. (*yaaru uuhisi)*
who-INDEF came who guess
'**Someone** (specific unknown) came. (*guess who)'

b. ***Ellig-aadaruu*** *hoogu.*
where-INDEF go
'Go **somewhere** (non-specific).'

(301) Quechua (Anchorena 1874, Cusihuáman 2001, Faller 2020, Haspelmath 1997:310, Parker 1976, Shimelman 2017, Weber 1989)

a. ***Ima-ta-chi*** *wambra yurapa waqta-n-chaw riqa-rqa-n.*
what-ACC-VAL boy tree behind-3SG-LOC see-PAST-3SG
'The boy saw **something** (specific unknown) behind the tree.'

b. ***Pi-cha*** *haqay-ta hamu-sha-n!*
who-CONJ over.there-ACC come-PROG-3
'**Someone** unknown is coming over there./I don't know who is coming over there.'

c. *Mana musya-pti-iki-qa* ***pi-ta-pis*** *tapuku-y*.
not know-CONV-2SG-TOP who-ACC-INDEF ask-IMPV
'If you don't know, ask **somebody** (non-specific).'

d. "*May-manta-n ka-nki-chis?*" *ni-spa* *tapu-wa-n* ***huk runa***.
where-ABL-BPG BE-2-PL say-nmlz.SS ask-1O-3 one man
'Where are you from? **a man** asked me.'

No indefinite pronouns

(302) Swahili (Hasplemath 1997: 302, Mpiranya 2015)

a. ***Mtu*** *a-li-ni-gusa*.
man 3SG-PAST-me-touch
'**Somebody** (specific unknown/specific known) has touched me.'

b. *Yesu a-li-mw-ambia kwamba a-wa-pe maskini* ***kitu***.
Jesus 3SG-PAST-him-tell that 3SG-them-give.SBJV poor thing
'Jesus told him to give **something** (non-specific) to the poor.'

(303) Irish (De Bhaldraithe 1959, Haspelmath 1997: 279, Windisch 1882)

a. *Tá* ***rud*** *agam le rá leat*.
is thing on.me for telling to.you
'I have **something** (specific known) to tell you.'

b. *Dúirt* ***duine*** *éigin liom é*.
told person certain to.me he
'**Somebody** (specific unknown) told me.'

c. *Abair* ***rud éigin***.
say.IMPV thing certain
'Say **something** (non-specific).'

(304) Filipino (native speakers: Christian Limbago and another anonymous speaker)

a. *Dalhan mo ako ng makakain!*
bring.IMPV you me of edible
'Bring me **something** to eat!'

b. *Meron tumawag pero di ko alam kung sino.*
there.is called but NEG I know if who
'**Someone** called, but I don't know who.'

c. *Meron akong ibibigay para sa iyo. Hulaan mo kung ano.*
have(there.is) I give.FUT for yours guess.IMPV you if what
'I have **something** to give you. Guess what.'