Two approaches for deriving a semantic map for indefinite pronouns

Bachelor thesis

Student: René van Gasteren, 3373509
Begeleider: Joost Zwarts
Cognitieve Kunstmatige Intelligentie
Geesteswetenschappen, Universiteit Utrecht
7,5 ECTS
30-06-2013

Contents

1 Introduction

The study of human language is one of the largest subjects within the bachelor Cognitive Artificial Intelligence (CAI). Our sophisticated use of language for communication and expression is an important quality that divides us humans from other animals and machines. CAI is more than a traditional AI or knowledge technology study; it includes many influences from both philosophy and cognitive science. It deals with the notion of intelligence and with the cognitive abilities of both humans and machines, including topics such as memory, reasoning, perception, action, knowledge, emotion and language.

1.1 Scientific field

Within this broad field of science human language can be studied in many different ways. Natural language processing can be described formally, for instance in categorial grammar, most of Chomsky's work, or statistical methods for language processing. Some research doesn't focus on understanding human language per se, but rather on making machines work with language: projects involving automatic translation and speech technology, or IBM's Watson. We have the field of formal semantics, which has a model-theoretic approach to language structure and meaning. A less formal approach is the field of psycholinguistics, which inquires how language is integrated in the human brain. It becomes a bit more philosophical when we want to know how we represent concepts, where meaning of linguistic units comes from, and what the relation is among language, other cognitive functions and the real world. We can also investigate language structure and language use, make cross-linguistic comparisons and investigate models to explain the variation and universal properties of languages.

This research falls under latest mentioned topic. The work of Haspelmath(1997) is taken as a starting point. The book studies indefinite pronouns (e.g. English *somebody*, *anybody*, Dutch *ergens*, *iemand*, or German *jemand*, *etwas*) from a typological perspective. Language typology is "the scientific study of variation and the limits to variation in the structure of languages" (Haspelmath, 1997: 7). There are thousands of different languages in the world, but all of them are used by the same beings in the same world in similar social situations. Natural languages can be seen as roughly equivalent (expressively speaking) instantiations of a more abstract notion of human language. Languages can both exhibit universal properties and unlimited forms of variation. The field of language typology tries to deal with this phenomena by using typological tools and methods to classify languages according to their uniformity and diversity.

Key to typological research is cross-linguistic investigation. Haspelmath(1997) used research from forty different languages as a base for his typological work. To represent his results he makes use of semantic maps. The semantic map method is a tool that has been applied successfully to other grammatical categories (Haspelmath, 2003). They are used for representing cross-linguistic variation of a linguistic unit or a group of linguistic units which express similar meaning or – as Haspelmath prefers to call it – have a similar function. This means that this research can also be seen as a topic within the field of lexical semantics. It involves studying the variation in word senses between ánd within languages. In conclusion, the semantic-map method is a tool (among many others e.g. the *list method* or *general-meaning method*, shortly discussed by Haspelmath(2003: 214)) for representing grammatical meaning.

1.2 Background information

Before we continue to the research question it is a good idea to have some background information about the thesis topic. We start with a rough notion of what the semantic map method involves. A simple example of a semantic map is one that can be made for kinship. Compare for instance English and Dutch. In Dutch, there is one word for both the son of your uncle and the son of your brother, while in English it is called *cousin* and *nephew* respectively. The same is the case for *cousin* and *niece*. A second difference is that English doesn't differentiate in gender between the son of your uncle and the daughter of your uncle. These findings can be represented in a semantic map. Figure 1 shows part of a possible semantic map for kinship [1]. We see the family relations (the functions) pictured as nodes in a graph and the words that express those relations (the lexical information) as areas on that graph. The functions are distributed based on our two findings above. When the functions are expressed by the same word they are connected by a edge or a path of edges.

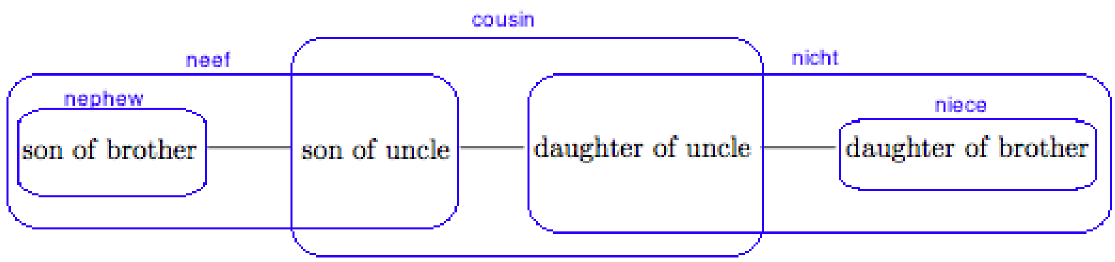

Figure 1: Part of a semantic map for kinship

Now that we understand a little bit of the way semantic maps work, let's introduce the one from Haspelmath(1997). Figure 2 is the conceptual space (i.e. a semantic map without the lexical information) of indefinite pronouns and Figure 3 is the semantic map for English. The functions will be discussed in detail in section 3.2.

[1]Instead of for instance *son of uncle* we might just as well used *son of aunt* or *son of parent's sibling*; it denotes the same thing, but we have to choose some name for the functions.

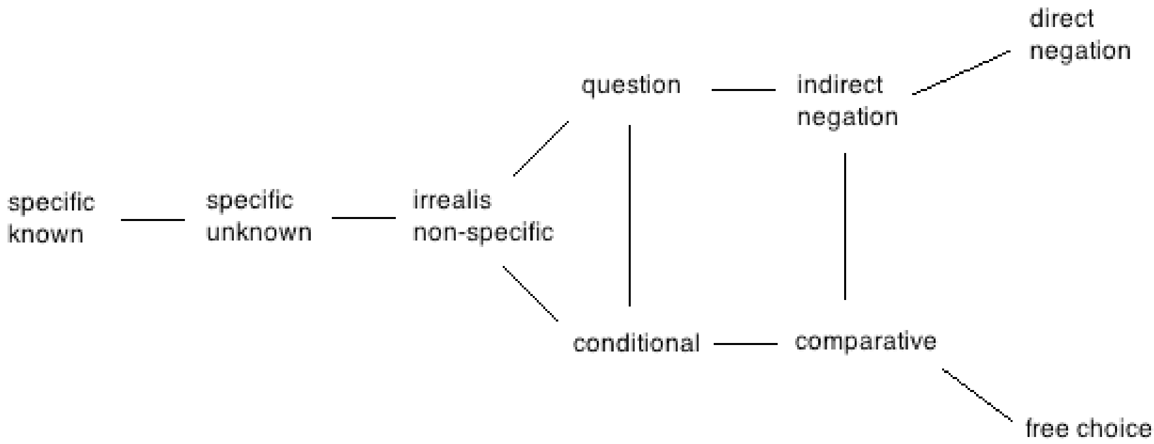

Figure 2: Conceptual space for indefinite pronouns from Haspelmath(1997)

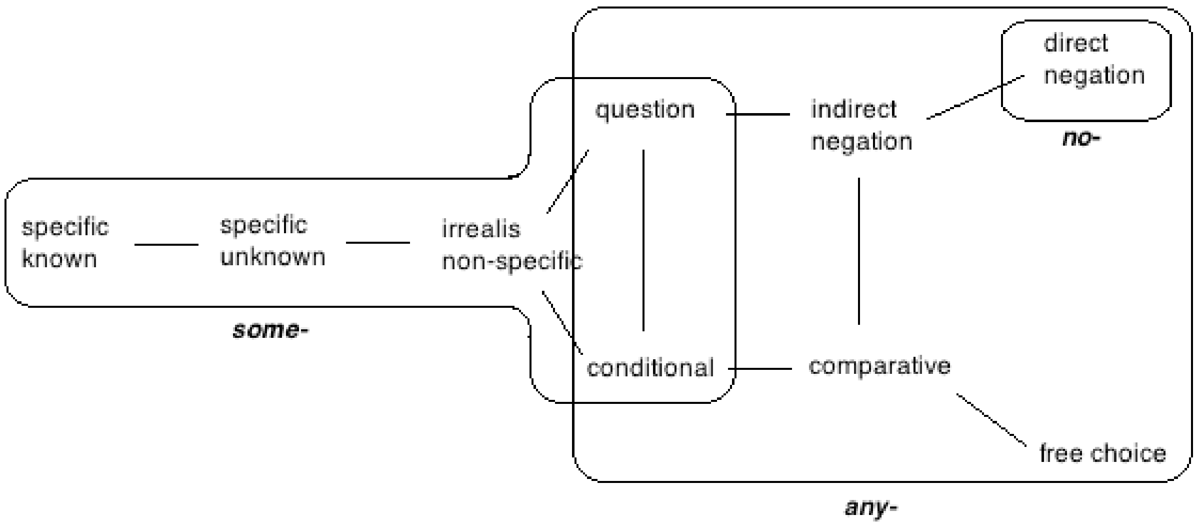

Figure 3: Semantic map for indefinite pronouns of English from Haspelmath(1997)

A crucial part of this thesis is the contrast between two approaches for deriving a conceptual space (discussed in Zwarts(2007)): a data-driven approach and a space-driven approach. The conceptual space in Figure 1 is derived from a data-driven perspective. The functions we choose in our little cross-linguistic inquiry about kinship were based on data from English and Dutch. It is based on how language is used. For example, we don't need to differentiated the functions *son of brother* and *son of uncle* for Dutch, because both are expressed by the same word. But in English *nephew* is used for *son of* *brother* and not for *son of uncle*, so a distinction based on English language use must be made. When we want to derive a conceptual space for kinship from a space-driven perspective, we semantically analyze the concepts of kinship; we ask ourselves what family relations are there and how are they different from each other in meaning. A possible conceptual space as a result using the space-driven perspective is represented in Figure 4.

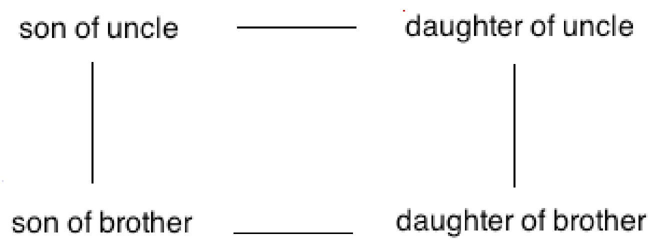

Figure 4: Part of a conceptual for kinship. The arrangement of the functions is derived by semantic analysis

Edges and path of edges are now related with similarity in meaning between the function (adjacent functions are semantically more alike). The biggest difference between the data-driven space (Figure 1) and the space-driven space (Figure 4) is the way *son of brother* and *daughter of brother* are arranged. It is semantically intuitive that these functions lie close to each other, but this is not justified by our data. Figure 5 shows that the lexical information fits the semantically derived space.

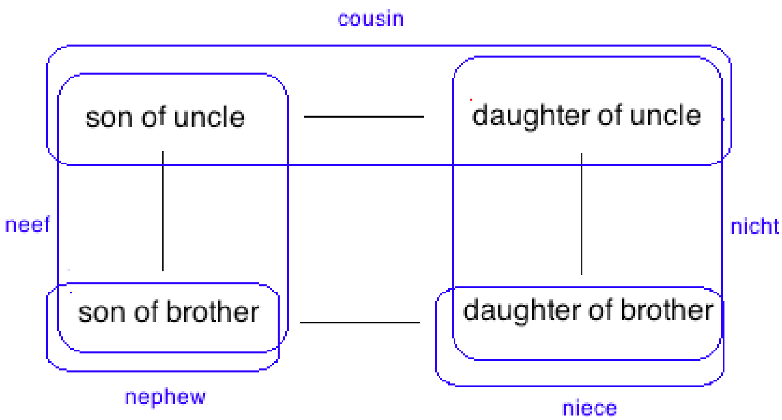

Figure 5: Part of a semantic map for kinship. The arrangement of the functions is derived by semantic analysis

Understand that when we say that a conceptual space is derived purely from a data-driven approach we do not mean that there is no semantic analysis involved. It is obvious that we need to have some sense of the notion, for instance, *son of brother* in order to conclude that *nephew* and *neef* share this function. What we mean is that the decision of which functions are included in the space and how they are arranged are purely based on data. An example can clarify this statement. Suppose that we apply a data-driven approach for kinship and we discover a language that uses different words for the son of an older brother and the son of a younger brother. From a data-driven perspective this justifies that we split the function *son of brother* in two separate functions. And, suppose there is no languages that have different words for the daughter of an older brother and the daughter of a younger brother, then there are no grounds for splitting *daughter of brother*. Now, imagine we apply a space-driven approach. When we have any purely semantic reason for applying both the function *son of oldest brother* and the function *son of youngest brother*, then on semantic grounds it seems natural to include the function *daughter of oldest brother* and *daughter of youngest brother* also. In conclusion, in a data-driven approach the data determines the inclusion and arrangement of functions in the space, while this is determined by semantic analysis in the matrix-based approach.

Haspelmath(1997) derives his conceptual space of indefinite pronouns from a data-driven perspective; he uses an empirical point of view based on cross-linguistic data. Every function and every connection is justified by his cross-linguistic set of data. One can ask now why is that these functions are arranged in this particular configuration? Or in other words, why do we use language in this particular way? How can we explain the conceptual space from a semantic or conceptual perspective? Haspelmath(1997) makes an attempt on answering these questions; he tries to demonstrate that functions that are semantically more similar are adjacent in the map, by characterizing the functions based on semantic features, i.e. defining each function by a list of binary distinctions. Let's use our kinship space of Figure 4 to exemplify this [2]. We might try to explain the conceptual space based on two features: *gender* and *generation*. The son of your uncle is a male (it has the property of being a male: M+), while the daughter of your uncle is not a male (M-), and the son of your uncle in of the same generation as you are (it has the property of being of the same generation: S+), while the son of your brother is from a later generation (S-). The number of edges between the functions in the space corresponds with the number of features the functions differ from each other, as can be seen in Figure 6. We have now given a feature-based semantic explanation for our conceptual space of kinship.

1.3 Research questions

In the end of section 1.2 it is mentioned that Haspelmath(1997) tried to explain his conceptual space for indefinite pronouns based on semantic features. He did not succeed; his explanation wasn't sufficient to explain the differentiation of every function in the

[2] Although this space in already derived by semantic analysis, I will use it anyway to explain the feature-based explanation. Besides, it is very plausible that this space can be derived data-driven when more languages are investigated.

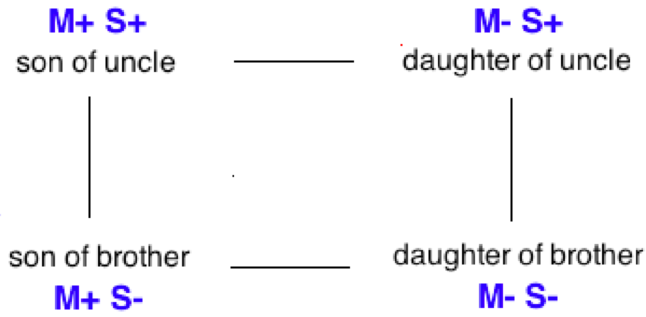

Figure 6: Feature based explanation for the conceptual space of kinship

space. The first research questions aims on improving Haspelmath's attempt. We will try to find a semantic description based on features to fully explain the configuration of the data-driven derived conceptual space of indefinite pronouns. For the second research question, we use the same features for a different purpose. Namely, we use this semantic description to derive a conceptual space from a space-driven approach. Next, we will compare this conceptual space with Haspelmath's conceptual space. This brings us to the formulation of the two research questions.

1) Can the arrangement of functions in the data-driven derived conceptual space of Haspelmath(1997) be explained according to the semantic feature method?

2) When a conceptual space is deduced from a space-driven perspective using the feature method, what are the similarities and differences with the empirically induced conceptual space of Haspelmath?

1.4 Structure of thesis

The structure of this thesis is based on the contrast of the two different perspectives for deriving a semantic map. In section 2, the most important concepts, namely semantic maps, functions and indefinite pronouns, are discussed in detail. Section 3 is an outline of the data-driven perspective of the work of Haspelmath(1997) and a detailed description of the functions of his semantic map. The data-driven perspective is contrasted with the semantic-based (space-driven) approach in section 4. The basis for a semantic explanation and a semantically deduced semantic map for indefinite pronouns is outlined here too. In Section 5 the results of the analysis in the previous sections are given and the the research questions are discussed. Section 6 concludes with a résumé and a general conclusion of the thesis.

2 Outline of essential concepts

In this section we describe three of the most important concepts for this thesis in more detail: semantic map, function and indefinite pronoun. Semantic maps were already introduced in section 1, but will be further explained here by giving a formal definition of what a semantic map is, and it is described how it is used for this research. Also a description of the notion of function is given. A brief description of indefinite pronouns concludes this section. Overall, this section is a description of the conceptual tool kit we will work with in this thesis.

2.1 Semantic maps

The semantic map method is a tool to describe and analyze patterns of multi-functionality of sense-related linguistic units or grammatical categories (e.g. tense, case, adpositions, indefinite pronouns). A semantic map typically consists of two parts (Zwarts, 2007): a conceptual space and a lexical matrix. A conceptual space is a geometrical representation of the functions and the way they are functionally related (i.e the connections between functions). A lexical matrix is some sort of list or table that shows for each linguistic unit which functions it can have. A semantic map is constructed by representing every word from the lexical matrix in the conceptual space. For example, part the lexical matrix for kinship would look like this:

	son of brother	son of uncle	daughter of uncle	daughter of brother
nephew	✓			
neef	✓	✓		
cousin		✓	✓	
. . .				

A conceptual space is seen as a universal model; all languages share the same space and it is because of this fact that cross-linguistic comparison is possible (Zwarts, 2007). A semantic map is often represented as an undirected graph with the functions as nodes. Similarity between the functions is expressed by adjacency of the nodes (i.e. having a common edge) (Haspelmath, 2003: 216). The length of the edges and the spatial orientation of nodes are insignificant and only for visualization purposes. A curved closed line indicates which functions can be expressed by a particular linguistic unit; it represents the lexical matrix. There are other methods similar to semantic maps (Zwarts(2007), Cysouw(2007)) or ways to modify them, letting them represent additional properties (Cysouw, 2007), but this is not relevant in the present work and we will not spend our time discussing them. Also we will not explore other purposes for semantic maps. Furthermore, we wont go into detail in ways of composing them or review the methodology, other than necessary. Lastly, a note on the term 'semantic map': it has some equivalents used in other literature, such as 'mental map', 'cognitive map' or 'implicational map' (Haspelmath 2003: 219). The terms 'mental map' and 'cognitive map' express the idea that the universal configuration of functions is a direct representation of a cognitive structure of meaning in the human mind. Whether this is true or not, I will stick to semantic maps only as a tool for linguistic representation.

2.2 Functions

We don't always make a purely semantic distinction between functions. In fact, Haspelmath (1997, 2003) makes very often use of a contextual or syntactic distinction as we will see in section 3.2. He prefers the term 'function', above the use of 'meaning', 'sense' or 'use' for two reasons. First, it is not always easy to see whether the difference is due to semantics or syntax (Look for instance at the following sentences inspired by a research discussed by Haspelmath(2003): *I eat soup with a spoon*, and *I eat soup with my mother*. Whether the contrast in use of 'with' is purely syntactic, or that there is a distinction in meaning, is open for discussion). The more neutral term 'function' is intended to cover both interpretations. Second, he avoids an indecisive and difficult discussion between different approaches within the field of lexical semantics. By using the term 'function' Haspelmath doesn't make any claim in favor of one of the approaches. Instead of 'function', I will sometimes also use to words 'use' or 'context' to refer to the functions in a semantic map.

When making a semantic map for a particular domain, functions need to be chosen and arranged. The way this is done depends on whether you want to derive them from data or based on semantic grounds. The arrangement of the functions is based on functional similarity between them. The functions must be arranged in such a way that every linguistic unit from the lexical matrix occupies a contiguous region in the conceptual space. In other words, the functions expressed by a lexical unit must form a connected subgraph of the conceptual space. This is called the contiguity principle. For example, the linguistic unit *neef* occupies a contiguous region in the semantic map of Figure 1. While a unit that only has the the functions *son of uncle* and *daughter of brother* does not occupy a contiguous region. When such a unit is found in a language the conceptual space must be adapted.

2.3 Indefinite pronouns

Everybody has an idea about what kind of words can be typed as indefinite pronouns. But if we are going to work with them and talk about them, we need to know exactly what we mean by indefinite pronouns. Since we work with cross-linguistically applicable concepts, the definition must be independent of language-particular properties. The boundaries between what is and what is not an indefinite pronoun are sometimes vague. Haspelmath(1997: 9-13) gives us an exact definition and explains the boundaries of his outline of indefinite pronouns. A complete understanding of this definition is not necessary for this research, so I only outline a very short version of Haspelmath's definition. Indefinite pronouns are grammatical elements that can replace noun phrases (e.g. something), but also adverbs or adjectives (e.g. somewhere, sometime, some kind of) and express an indefinite reference. He explicitly excludes other categories of pronouns (personal, demonstrative, relative, and interrogative pronouns, and several other pronouns that are sometimes categorized as indefinite pronouns).

3 Haspelmath's work

Now that we roughly understand the the work of Haspelmath(1997), and have an idea of the goal of this thesis, we are going to look at the process of Haspelmath's typological study in more detail. As mentioned earlier a semantic map can be derived from data or by semantic analysis (or using both; many studies derive semantic maps from an interaction between the two perspectives). To make a clear contrast, Haspelmath empirical data-driven study is discussed in this section and semantic analyses are discussed in section 4. The biggest part of this section is an analysis of the functions used by Haspelmath. Such an comprehensive outline is necessary to understand Haspelmath's work precisely and helps us to prepare for the the discussion in section 4 and 5.

3.1 Typological study

Haspelmath(1997) divides a typological study into four steps. We can apply this general plan to our investigation on semantic maps for indefinite pronouns. Step one, formulate a cross-linguistic definition of indefinite pronouns. Two, provide a complete classification of the typological variation of indefinite pronouns. Three, represent the classifications and the cross-linguistic variation using the semantic map-method. Four, give an explanation for the semantic map.

We already described the first step of Haspelmath's research in section 2.3. The second step is the analysis and categorization of the functions expressed by indefinite pronouns, and is described in section 3.2. The third step involve the the selection and arrangement of functions. Haspelmath(1997: 61) describes the principle for function selection: "Whenever two roughly comparable categories in two languages turn out to differ in one type of environment or meaning, this is sufficient for setting up a separate use." To understand what this means, recall the kinship map. The categories *nephew* and *neef* are used to denote *son of brother*, but *neef* denotes also *son of uncle*. In order to differentiate between *nephew* and *neef*, a function *son of uncle* is introduced. We're not discussing function selection any further in this research. For function arrangement there is no standard procedure (Haspelmath,2003: 217), only a criterion that the arrangement must be in a way that all multifunctional indefinite pronouns occupy a contiguous area in the semantic map. The first three steps leads to the conceptual space as seen in Figure 2 and to a semantic map for all 40 languages, like the one in Figure 3 for English. The fourth step is giving an (semantic) explanation for the observed patterns; why are the functions in the semantic map for indefinite pronouns arranged in the way induced from cross-linguistic data? This question is similar to our first research questions and will be outlined in the rest of this thesis and answered in section 5.

3.2 Outline of the functions

Here, we discuss all the functions selected by Haspelmath(1997). To avoid redundant definitions, we will divide the total set of function in subsets, each discussed in a different subsection. Haspelmath(1997) distinguishes nine functions for his semantic map on indefinite pronouns. However, more distinctions can be made (Aloni et al., 2010); by comparing more different linguistic items semantically or cross-linguistically in detail, we can modify the conceptual space and add three more uses [3]. The extended conceptual space is shown in Figure 7. [4] In this subsection all twelve functions are given a description and exemplification to motivate the functional distinction expressed by indefinite pronouns. All examples will be in English, but remember that the functions are universal.

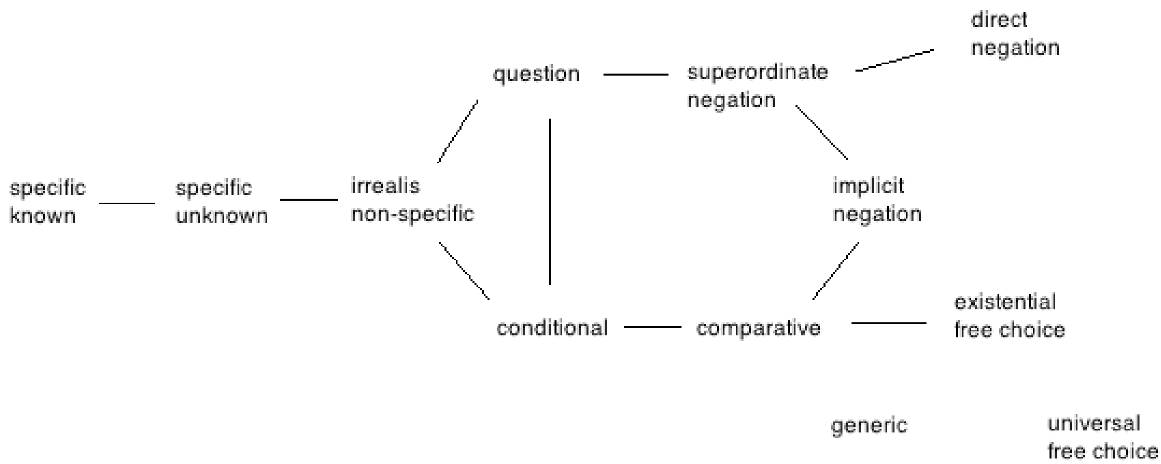

Figure 7: Conceptual space of indefinite pronouns based on (Aloni et al., 2010)

3.2.1 Negation

Negative indefinite pronouns are used in negative sentences where the scope of negation extends over the indefinite (Haspelmath, 1997).

(1) I will not say anything.

(2) Mary doesn't believe that anyone did their homework.

(3) My keys are nowhere to be found.

(4) Nobody left the house yesterday.

(5) The security denied giving access to anyone.

Notice that the English no-series (3-4) already have a negative morpheme itself, while other indefinite pronouns (1-2) co-exist with a verbal negation. (5) is an example of a

[3]It is actually not adding three functions per se, but splitting two existing functions into sub-functions.

[4]There are a few differences compared to Haspelmath's conceptual space. The function *indirect negation* is here split into *superordinate negation* and *implicit negation*. *Free choice* is replaced by *existential free choice*, and *generic* and *universal free choice* are added to the space.

weaker form of negation.

The function of negation can be subdivided into three of our functions: *direct negation*, *superordinate negation* and *implicit negation*. The latter two can be treated together because the indefinite pronouns used in both contexts are generally the same (Haspelmath, 1997, 33). But based on semantic grounds a distinction can be made (Aloni et al 2010: 3).

Direct negation
The function of direct negation is given to indefinite pronouns used as an argument of a negated clause. The use of the no-series (6-8) is exclusively for *direct negation* (Haspelmath, 1997).

(6) Nobody came to the party.

(7) I noticed nothing.

(8) The cowboy was nowhere to be found.

(9) I can't find it anywhere.

(10) I have not seen any of them.

Superordinate negation
When an indefinite pronoun appears in a subordinate clause of a negative sentence, it functions as a *superordinate negation*. This function is equivalent to the function called *anti-morphic* by Aloni et al.(2010).

(11) I don't think that anybody come.

(12) John hasn't said that anything is wrong.

(13) I don't understand why anyone would do that.

Implicit negation
An indefinite pronoun used as *implicit negation* appears in a subordinate clause of a sentence with an implicitly negative expression like *without*, *deny*, or *refuse*. Aloni et al.(2010) named this function *anti-additive*.

(14) The company refuses to make any decision.

(15) The security will deny access to anybody.

(16) The man avoided looking anyone in the eyes.

(17) Only Mira has anything substantial to report.

3.2.2 Negative polarity

A negative polarity item is an expression that needs to be in the scope of a monotone decreasing element (e.g. negation (18), other negative environments (19-21), interrogative (22), conditional (23) or comparative (24)).

(18) He didn't do anything to help her.

(19) I doubt anyone can find the entrance.

(20) Few people show any interest in global issues.

(21) Without saying anything, she ate the soup.

(22) Is there anything new?

(23) If you tell anybody, well punish you.

(24) An apple is healthier than a brownie.

Negative contexts, discussed in section 3.2.1, are a subdivision of the negative polarity environment.

Question

When we talk about questions as a function of indefinite pronouns, we only mean polar questions (i.e. a yes–no question). I believe the category of *question* doesn't need any further explanation.

(25) Did you see anything?

(26) Can I work anywhere in the building?

(27) Can you give something to him?

Conditional

An indefinite pronoun is given a *conditional* function, when it appears in the antecedent of a conditional expression (Haspelmath, 1997).

(28) If you see anybody, tell me immediately.

(29) If you go anywhere, send me a card.

(30) Everyone who likes any kind of seafood will like our new seafood sticks.

Of course, an indefinite pronoun can also occur in the consequent of an conditional statement, but an indefinite pronoun only has the function of conditional when it appears in the scope of the if-statement. For example, (31) has a *free choice* use (discussed in 3.2.4) and (32) is in the context of *irrealis non-specific* (discussed in 3.2.3).

(31) If you own a travel card, you can travel with any transport you like. [5]

(32) If you find a drugstore, buy me any painkiller.

Comparitive

Indefinite pronouns used in a comparative context have the function *comparative* (Haspelmath, 1997). But this is only the case when the pronoun occurs on the right side of 'than'.

[5] Indefinite determiners such as 'any' in example (31-32) or 'some' are included in the study (Haspelmath, 1997: 11)

(33) The boy can run faster than anyone in his class.

(34) This place is better than anywhere else.

(35) This lunch is as good as any food in the cafeteria.

(36) ? Anyone runs faster than the boy.

(37) ? Anything is better than this.

3.2.3 Specifity and non–specifity, known and unknowm

Some languages have indefinite pronouns that are exclusively used for either specific use or non-specific use (Haspelmath, 1997). Haspelmath(1997: 38) says: "I will say (preliminarily) that an expression is specific if the speaker presupposes the existence and unique identifiably of its referent." In other words, when there exist something the speaker can refer to, by using for example a pronoun or an definite determiner in a subsequent sentence, the indefinite pronoun is specific. The question whether a speaker is able to identify the referent (i.e. knows the referent) of the indefinite pronoun presupposes that the indefinite pronoun is specific. For non-specific meaning the indefinite pronoun is necessarily unknown (Haspelmath, 1997: 45). For English it is often the case that the context must be used to infer whether a specific or non-specific, or a known or unknown function is expressed:

(38) a. James wants to marry some$_{non-specific}$ greek woman, because he believes the most beautiful women are greek.

b. James wants to marry some$_{specific}$ greek woman. He met her on vacation last summer.

(39) a. Suzy met with someone$_{unknown}$ near the bar, I couldn't figure out who is was.

b. Suzy met with someone$_{known}$ known to me near the bar.

In some contexts only specific indefinites are allowed (Haspelmath, 1997: 39): "Such is the case, in particular, in affirmative declarative sentences in the perfective past or in the ongoing present. In such prototypical realis sentences, the speaker is committed to the existence and identifiability of the entity and indefinites of the non-specific series are simply unacceptable". Indefinite pronouns in a realis sentence can only have the function specific known or specific unknown. A specific indefinite pronoun can either be known or unknown.

Specific known

In English the difference between known and unknown isn't in the semantics of the indefinite pronouns itself, but needs to be derived from the pragmatics. Some languages have different indefinite pronoun for the two different specific indefinite pronoun, English has not (Haspelmath, 1997).

(40) She wanted to acquire some object, so I gave it to him.

(41) On Saturday we will go somewhere. I know where, but I'm not allowed to tell.

Specific unknown

(42) Somebody came, but it wasn't my mother

(43) Ivan wants to sing some romance, you have to ask him what he wanted precisely

In all other contexts, except *specific known* or *specific unknown*, the indefinite pronoun has a non–specific reading (Haspelmath, 1997).

Irrealis non–specific
An irrealis context describes an event that hasn't been realised yet. Irrealis indefinite pronouns appear for example in imperatives, future tense, sentences expressing uncertainty or sentences with 'want'. Although an irrealis reading of an indefinite pronoun is not necessarily non–specific, this function only accounts for non–specific irrealis (Haspelmath, 1997).

(44) Visit me sometime.

(45) Apparently someone complained about her.

(46) Someone will pay for the damage made!

In the literature there is no agreement on the specification of the term 'irrealis'. By some definition questions and free choice readings of indefinite pronouns are also irrealis. Haspelmath(1997) never mentions this issue, and because no problems arise, we stick with the definitions that *irrealis non-specific* occurs only in imperatives, future tense, sentences expressing uncertainty or sentences with 'want'.

3.2.4 Free choice

Free choice indefinites are semantically similar, but not the same, to universal quantifiers (Haspelmath, 1997: 48).

(47) You can pick any apple.

(48) You can pick every apple.

Aloni et al. (2010) adds two more uses – *generic* and *universal free choice* – to the semantic map that are related to the free choice function, but she doesn't make clear how this relation is supposed to be interpreted. Haspelmath already mentions the generic context as a subset of *free choice*, so it makes sense to think that we can divide Haspelmath's *free choice* into at least two sub-functions: the free choice (49) of Aloni et al.(2010) (which we already called *existential free choice* in Figure 7) and *generic* (50).

(49) You may kiss anybody.

(50) Any cat is an animal.

Existential free choice

The most typical kind of environment in which an existential free choice indefinite pronoun appears are sentences that express possibility (Haspelmath, 1997).

(51) I can choose anything.

(52) We allow you to buy anything within the budget.

Furthermore *existential free choice* indefinite pronouns are permitted in hypothetical (53) or counterfactual contexts (54), and in contexts expressing sufficient conditions (55), independently of tense or modality. They are not allowed in context where non-specific indefinites are not allowed, i.e perfective past (56) and ongoing present contexts and contexts of necessity (57). Free choice indefinite pronouns in future tense (58), imperatives, and contexts with 'want' (59) are just a bit awkward (Haspelmath, 1997).

(53) I would give anything to see that.

(54) At that point, I think I would have accepted anything.

(55) Any amount is adequate.

(56) * She bought anything.

(57) * Any amount is necessary.

(58) ? Next week, we will go anywhere.

(59) ? Jane want to read any book in the library.

Universal free choice

Haspelmath(1997) never mentioned a type of function that Aloni et al.(2010) calls *universal free choice*. But from the sentence example (60) of the latter we can derive some information about this use of this context type.

(60) John kissed any woman with red hair.

Whereas Haspelmath's free choice is similar to a universal quantifier, *universal free choice* is in fact equivalent to a universal quantifier.

(61) Jane eats any biscuit she gets her eyes on.

(62) Sam says anything that pops into his mind.

In his definition for indefinite pronouns Haspelmath explicitly excludes universal quantifiers. Since indefinite pronouns with a universal free choice function are equivalent with universal quantifiers, I will exclude this function from our research.

Generic

Haspelmath(1997: 50) already mentioned generic use of an indefinite pronoun. It functions in a pretty straightforward way.

(63) Any dog has four legs

(64) Any cat is a mammal

3.3 Analysis of functions

To fully understand the notion of function we have to go further into detail about the question what it means for an indefinite pronoun to have or have not a certain function. This section should clear all issues that arise and questions that came up when investigating the functions.

Aloni et al. (2010) states: "In order for an indefinite to qualify for a function, it must (i) be grammatical in the context of the function; and (ii) have the semantics that the function specifies." The functional description seems to license the use of only particular indefinite pronouns in a certain context or environment.

The first criterion is easy to understand: it excludes uses of indefinite pronouns that compose ungrammatical sentences, such as (65-67).

(65) *Anybody called.

(66) *You must marry anybody.

(67) *Anyone lived in a pretty town.

The second criterion states that the indefinite pronoun must have the meaning specified by its function. Consider (68-70). (68-70)a don't seem to be consistent with the semantic map of Figure 3. For instance, according to the semantic map *nobody* can only have the function of *direct negation* and not of *question*. Why can *nobody* appear in an interrogative environment, without having the function of *question*? Because *nobody* doesn't have the meaning that the function of question specifies, thereby not satisfying criterion (ii). (68-70)b do satisfy both criteria.

(68) a. Did nobody came to the party?
b. Did anybody came to the party?

(69) a. The boy runs faster than someone in his class.
b. The boy runs faster than anyone in his class.

(70) a. You can find it somewhere.
b. You can find it anywhere.

Only the meanings of the indefinite pronouns in (68-70)b is specified by the function. The environment of *question* (68) *comparative* (69) and *free choice*(70) seem to license only the any-series (and in case of *question* also the some-series). Unfortunately, both Haspelmath(1997) and Aloni et al.(2010) neglect to give an exact specification of the functional description other than by exemplification. So why are other series excluded here? Is it only because it lacks the semantics the function specifies? Maybe, but we can come up with another explanation that can be illustrated by the following examples.

(71) a. I don't think that somebody goes to the party.
b. Somebody goes to the party.

(72) a. I don't think that anybody goes to the party.

b. *Anybody goes to the party.

The indefinite pronoun in (71)b is licensed by a specific (un)known environment and it stays licensed in this way when it appears in (71)a. It doesn't matter whether the sentence becomes a subordinate negation sentence afterward, the indefinite is still licensed by the specific (un)known environment. In the case of (72)b; this sentence is ungrammatical, so the indefinite pronoun is not licensed by anything yet. This occurs when the environment changes, for example into a superordinate negation environment of (72)a. The fact that an indefinite pronoun can occur in a specific context, does not necessarily mean that the context license the use of that indefinite pronoun. The same explanation can be applied to (68-70)

The criteria of Aloni et al. (2010) and concept of licensing helps us also to understand why some indefinite pronouns seem to have multiple functions. Consider (73)a; the indefinite pronoun is used in a negative context, but also seems satisfy the criteria for *free choice* (which we will see is does not).

(73) a. I can't find it anywhere

b. I can find it anywhere

I believe we are deceived by the fact that *anywhere* is multifunctional by itself in English. In the context of (73)a it only has an negated use. This can be shown by looking at another language, in this case Dutch:

(74) Je kan dit product **waar-dan-ook** krijgen.
You can this product anywhere get.INF
'You can get this product anywhere'

(75) *Je kan dit product niet **waar-dan-ook** krijgen.
You can this product not anywhere get.INF
'You can't get this product anywhere'

The indefinite pronoun 'waar dan ook' in (74) has a free choice function. When we use that same indefinite in a the scope of negation, we get a grammatically incorrect sentence (75), because 'waar dan ook' can't have a negated function. *Anywhere* in (73)a has the function of negation, and not a free choice function, while it has the function of *free choice* in (73)b. The same explanation can be used for (76-79) (examples taken from Haspelmath (1997)).

German:

(76) Hast du **etwas** gesehen?
Did you see anything$_{question}$?

(77) *Das steht dir besser als **etwas** anderes
*This suits you better than anything$_{question}$ else.

Lithuanian:

(78) Aplanky-kite mana **kada nors**.
visit-IMPV.2PL me when INDEF.
'Visit me sometime$_{non-specific}$'

(79) *Aplanky-kite mana **kaž-kada**.
visit-IMPV.2PL me INDEF-when.
'Visit me sometime$_{specific}$'

The indefinite pronoun *etwas* in (76-77) is licensed by interrogative environments, whereas the comparative environment does not license the use of *etwas.* Both the question and comparative environment license the use of *anywhere.* The same is the case for (78-79): the imperative environment licenses the use of a non-specific indefinite pronoun, but not a specific pronoun. This explains why some indefinite pronouns seem to have more than one function; that is only because the words can be used in different environments. However, only one function is expressed when used in a particular context.

We can conclude our in-depth analysis of functions with the following statements:
1) An indefinite pronoun has one or more functions
2) A function can be expressed by any number of indefinite pronouns
3) An indefinite pronoun used in context has only one function
4) A function has a functional definition, i.e. a definition based on syntax, semantics and pragmatics.
5) An environment (context or sentence) licenses only one function.
6) A function can be defined by the environment it is licensed by.

The final conceptual space for our analysis is shown in Figure 8. *Universal free choice* is disposed and *generic* has been given a place close to *existential free choice.*

4 Alternative approach

The data-driven approach used by Haspelmath is most common in semantic mapping (Zwarts, 2007). But what we are also interested in for this research is whether it is possible to deduce a conceptual space based on semantic grounds. The semantic theory we will use for the investigation is, as mentioned before, the semantic feature method. A differentiation of functions is made based on a checklist of features. In this section we will discuss this method in detail.

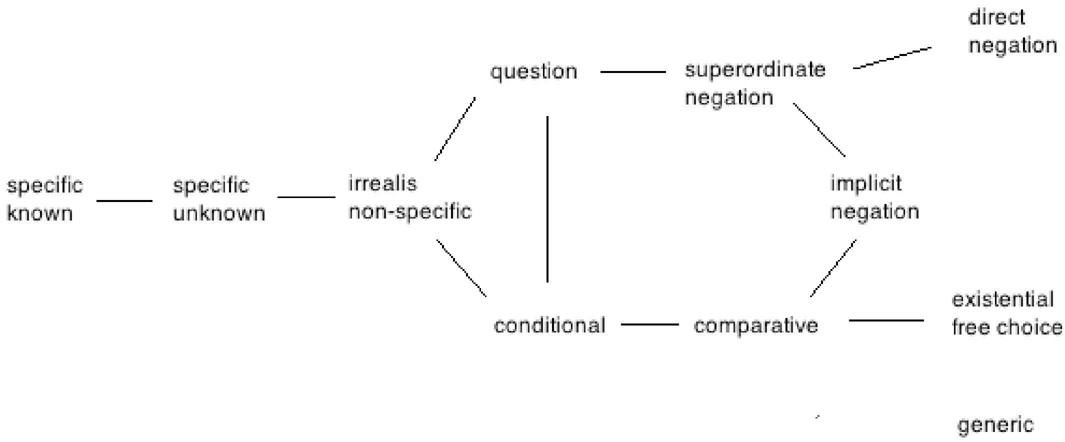

Figure 8: Conceptual space of indefinite pronouns used in this research

4.1 Contiguity and classical categorization

According to Zwarts(2007), using the data-driven method, you make no assumptions about how the different functions relate to each other. The conceptual space is induced from data, but why is it that some functions are closer to each other, while others are further apart? Another important question is why some set of functions form a contiguous area, while other sets don't. The only principle behind contiguity seems to be some sort of family resemblance between the functions. Zwarts wonders whether all possible contiguous regions can also be defined in a semantic way, in particular the classical way, i.e. categorization based on a list of necessary-and-sufficient features or a checklist of essential features. Does the contiguity principle tell us anything more than what can also be said in terms of features, i.e. can the family resemblance seen in the conceptual space be completely explained in a classical way? He suggest that such conceptual spaces do exist and gives a motivation for his claim. This brings us a little further in answering the first research question. There are conceptual spaces that can be completely defined from a classical point of view. For further details about this topic, I refer to Zwarts(2007). Now, we need to find out whether the space of indefinite pronouns is one of them.

Haspelmath(1997: 119-121) leads us in the right direction with his explanation of the semantic map based on features. He divides the conceptual space by characterizing the functions according to five binary features (Figure 12-16 in appendix B). (Note that this is roughly the same way we divided the description of function in section 3.2.) However, this method still has some shortcomings: it doesn't explain the difference between the functions *comparative*, *conditional* and *comparative*, and the distinction between *superordinate negation*, *implicit negation* and *direct negation* is undefined too. Haspelmath notes the first shortcoming and adds an ad hoc non-classical explanation to solve this, but he forgets the solve the second shortcoming (Haspelmath, 1997: 121-122).

In order to answer the first research question we can use Haspelmath's attempt, but we need to improve this. More features are needed to semantically differentiate every function from each other. To identify the necessary features we use the work of Aloni et al.(2010). In their research they developed a binary branching decision tree assigning each instance of an indefinite pronoun to exactly one function. The decisions or tests are based on the semantic feature method. In the next section the features used by Aloni et al.(2010) are discussed.

4.2 Features

At first the difference between functions and features will seem a bit arbitrary. Sometimes, the same characteristic is used for both a (part of a) function and a feature. For example the feature *specific* is used to describe the functions *specific known* and *specific unknown*. Or even more confusing is the use of *question* or *comparative* as a function and a feature. To work with both notions it is necessary to understand the difference in nature. Features are properties of a function, and thereby indirectly also characteristics of an indefinite pronoun. An indefinite pronoun can (see section 3.3) express only one function, but a function can have several features. The features make up the difference between functions. For example the feature *specific* is exactly the property that describes the distinction between the functions *specific known* and *specific unknown* from all other functions. By means of the feature method every function can have its unique set of properties.

Two things have to be said before we start. First, the tests are very generally and briefly defined, which might give some difficulties when we look at them in more detail, but we will see that we will overcome those. Second, Aloni et al.(2010) make use of an operator in the test without explaining what it actually is and how it should be used in the test. So it's a good idea we define the notion of an operator first, based on the way its used by Aloni et al.(2010).

In most tests an operator **Op** is used. The operator seems to replace all the words, except for the indefinite pronoun in the sentence. However, this is too simplistic, as we will see when we describe the tests in more detail. An operator is the context or environment in which an indefinite pronoun appears and is invariant of the exact words used. Part of the sentence is replaced by the operator, which is in fact just an abstract form of the context in which it appears. Examples of types of contexts:
- modality (can, may, must, will, would etc.)
- tense/aspect (past, present, future)
- mood (interrogative, conditional, irrealis, imperative)
- negation, generic.

The plan for this section is like this: I will introduce the tests one by one and extend them using the information we have so far and modify them to get rid of the upcoming problems.

Specificity

Test: **Sentence**: ... indefinite$_i$ ... **Possible Continuation**:... pronoun$_i$ ... [S+]
This test is based on whether there is actually an entity to refer to, using an indefinite pronoun. If in a possible next sentence we can refer to that entity using a pronoun, then the function of the indefinite pronoun is characterized as specific. This is equivalent with the description given in section 3.2.3, where we defined the function *specific (un)known*.

(80) She wants to discuss something$_i$ with you, it$_i$ is very important. [S+]

(81) Somebody$_i$ want to see you. You can meet him$_i$ by the entrance. [S+]

(82) John didn't see anybody. * He is very tall. [S–]

(83) Did you hear anything in the attic? * It is very loud. [S–]

As said earlier, all contexts but specific known and specific unknown are marked as non-specific. But how can it be then that the following examples pass the test?

(84) If you tell anybody$_i$, we'll punish you and him$_i$ too. [S+]

(85) Only Mira has anything$_i$ substantial to report. It$_i$ ought to be very relevant. [S+]

(86) The boy can run faster than anyone$_i$ in his class. They$_i$ are all in a wheelchair. [S+]

This is very well known phenomena in the literature, that result into false positives for the specificity test. We excluded those and won't pay any further attention to it.

Known

Test: **Sentence**: ... indefinite$_i$ ... **Possible Continuation**: Guess interrogative$_i$?
By using the *Guess who/what/where/...* phrase, the test specifies whether the speaker has knowledge about the identity of the entity referred to by the indefinite pronoun.

(87) Look, somebody$_i$ is running. Guess who$_i$? [K+]

(88) Did you see anything$_i$? Guess what? [K–]

(89) Steven can be anywhere$_i$ right now. Guess where$_i$? [K–]

There a no relevant problems with the definition of the known feature.

Universal meaning

Test: ... **Op** (... indefinite ...) $\Rightarrow$ ... $\forall x$ (**Op** ... x) ...
This test distinguishes a universal meaning from an existential meaning.

(90) Visit me sometime $\not\Rightarrow$ for every time x: visit me at x. [$\forall$–]

(91) David has a faster car than anybody in his family $\Rightarrow$ for every car x in the family of David: David's car is faster than x. [$\forall$+]

(92) He left without saying anything $\Rightarrow$ for every x: he left without saying x. [$\forall$+]

The characteristic of universal meaning is never mentioned by Haspelmath, but it's a reliable feature which give us no problems at all.

Polar question

Aloni et al.(2010) introduce this test only for pragmatic reasons, to distinguish the context of questions from the context of irrealis non–specific. The only thing the test tell us, is that a question is a question and everything else is not a question. So this is not a very valuable feature, moreover if we can find another feature which differentiate questions from irrealis non–specific use, the feature of polar question is obsolete.

Anti-additivity

Test: $\mathbf{Op}\ (a \vee b) \equiv \mathbf{Op}(a) \wedge\ \mathbf{Op}(b)$

Anti-additivity is a term used in set theory. It is a logical property of a function and can be best understood by the following definition (van der Wouden 1994: 32):

Let A and B be two Boolean algebras. A function from A to B is anti–additive iff for arbitrary elements $X, Y \in A : f(X \cup Y) = f(X) \cap f(Y)$.

The function f in this definition is replaced by an operator, and the equivalence relation replaced by entailment by Aloni et al.(2010) but this doesn't change anything for us.

(93) I can't find it anywhere. [I can't find it in the basement or in the attic $\equiv$ I can't find it in the basement and I can't find it in the attic] [A+]

(94) If you tell anybody, we'll punish you. [If you tell John or Mary, we'll punish you $\equiv$ If you tell John, we'll punish you and If you tell Mary, we'll punish you] [A+]

(95) Chris met someone at the bar. [Chris met John or Mary at the bar $\not\equiv$ Chris met John at the bar and Chris met Mary at the bar] [A–]

(96) Go and find some sailor. [Go and find John or Mary $\not\equiv$ Go and find John and Go and find Mary] [A–]

Genericity

Test: ... indefinite ... $\equiv$ ... plain generic indefinite ... [G+]

The test for genericity is a pragmatic one: it differentiate the two functions it is supposed to in the decision tree, but it does not distinguish generic from non–generic interpretations in general. Take the following example:

(97) Someone went to the store ≡ an X went to the store.

When you interpreted *someone* as known or unknown the equivalence relation seems to hold, but the sentence can by no means be read as a generic sentence. The same thing can be said for some other functions.

(98) He didn't do anything to help her. ∼ He didn't do a thing to help her.

(99) The boy runs as fast as anyone in his class. ∼ The boy runs as fast as a person in his class.

(100) Did you see anything? ∼ Did you see a thing?

(101) Visit me sometime. ∼ Visit me a time.

(102) You may kiss anybody. ∼ You may kiss an X.

The test for genericity is unusable for this purpose. It seems safe to consider only generic contexts as generic, and other functions as non-generic, giving the feature the same status as the question feature.

Negation

Test: $\mathbf{Op}(a \vee \neg a)$ is inconsistent ... [N+]

The test for negation is easy to understand and correct. A sentence of the form $a \vee \neg a$ is always true and when such a sentence is negated, as is the case in negative context, the form is $\neg(a \vee \neg a)$ and inconsistent.

Anti-multiplicativity

Test: $\mathbf{Op}\ (a) \vee \mathbf{Op}(b) \equiv \mathbf{Op}\ (a \wedge b)$ ... [M+]

Anti-multiplicativity is another logical property of functions and can be defined as followed (Van der Wouden, 1994: 32):

Let A and B be two Boolean algebras. A function from A to B is anti-multiplicative iff for arbitrary elements $X, Y \in A : f(X \cap Y) = f(X) \cup f(Y)$.

(103) John didnt see anybody. [John didnt see Mary or John didnt see Sue ≡ John didnt see (Mary and Sue)] [M+]

(104) The bank avoided taking any decision. [The bank avoided taking decision A or the bank avoided taking decision B ≢ The bank avoided taking (decision A and decision B)] [M–]

(105) The thief escaped without anything. [The thief escaped without money and jewels ≢ The thief escaped without money or the thief escaped without any jewels.] [M–]

Clausal negation

The test for clausal negation seems interesting to differentiate negative contexts. *Direct negation* (106-107) is clausal negation, but *superordinate negation* (108) and *implicit* *negation* (109) is not.

(106) Nobody failed the test. [D+]

(107) We haven't touched anything in the museum. [D+]

(108) I don't believe she is anywhere in the building right now [D–]

(109) I refuse to do anyone's dirty work [D–]

Free choice

... $\mathbf{Op}(a \vee \neg a)$ is informative [F+]
Unfortunanly, Aloni et al.(2010) don't explained the relation between passing the test and having the property of free choice. It still is a usefull test to distinguish *existential free choice* and *generic* from the other functions

(110) You can kiss anybody. [You can stay or go] [F+]

(111) Any cat has a soft skin. [Any cat or no cat have a soft skin] [F+]

(112) Can anybody help me carrying this box. [Can somebody or nobody help me] [F-]

(113) Ivan wants to sing some romance. [Ivan wants to stay of go] [F-]

Comparative construction

The test for comparative construction has a similar purpose as the polar question test. It only distinguished the only comparative function from the rest.

4.3 Justification for the semantic features

Some features are more useful than others. *Polar question*, *genericity*, *comparative construction* and *clausal negation* only distinguishes one functions from the rest. Honestly, using features like these, feels a bit like cheating. The only thing the feature *polar question* does, is telling us that the function *question* is a question and all the other function are not. The same is the case for the other three. Moreover, Aloni et al. (2010) only seems to introduce this features, or in their case tests, to differentiate between two function in a last step in the decision tree. As we will see *polar question* and *genericity* won't be necessary to divide respectively *question* and *generic* from other functions. Keeping *comparative construction* and *clausal negation* is, unfortunately, necessary for respectively dividing *comparative* and *conditional*, and *direct negation* and *superordinate negation* from one another.

Preferably, we would like to have a list of simple, indefinable and universal features, for our explanation of the conceptual space, just like the semantic primitives of the Natural Semantic Metalanguage described in Goddard(2010). But this is a very time consuming process. Moreover, who can tell whether a feature is actually a fundamental property? We are not claiming that our list of features is the correct one. Whether we can of cannot explain the conceptual spaces with the use of semantic features highly depends one the features used for the explanation. Therefore, the choice for a set of features is mostly a pragmatic one. As we will see, our semantic feature explanation covers most of the data from Haspelmath's research and our analyses. So we might say that our choices for features is justified for the most part. If we want to completely formulate the correspondence between the data and the semantic features, we either have to adapt the list of feature, ór find new data.

Another remark on the notion of semantic features that must be mentioned, is that not all features our purely semantic or logical. The feature *Comparative construction* for example is rather a contextual of syntactical property. It shows that the notion of features is richer than just semantic properties, or that we also need syntactical (or pragmatic) properties to fully characterize the functions.

Table 1: Features values per function

Test \ Functions	SK	SU	IR	Q	CA	CO	DN	SN	IN	FC	GEN
Specificity	S+	S+	S-	S-	S-	S-	S-	S-	S-	S-	S-
Known	K+	K-	K-	K-	K-	K-	K-	K-	K-	K-	K–
Universal meaning	∀-	∀-	∀-	∀-	∀+	∀+	∀+	∀+	∀+	∀+	∀+
Anti-additivity	A–	A–	A–	A+	A+	A+	A+	A+	A+	A+	A-
Negative meaning	N–	N–	N–	N–	N-	N-	N+	N+	N+	N-	N–
Anti-multiplicativity	M–	M–	M–	M–	M–	M–	M+	M+	M-	M–	M–
Clausal negation	D–	D–	D–	D–	D–	D–	D+	D–	D–	D–	D–
Comparative Con.	C–	C–	C–	C–	C–	C+	C–	C–	C–	C–	C–
Free choice	F–	F–	F–	F–	F–	F–	F–	F–	F–	F+	F+

Table 2: The number of different feature values between functions

	SK	SU	IR	Q	CA	CO	DN	SN	IN	FC	GEN
SK	×	1	2	3	4	5	7	6	5	5	4
SU	×	×	1	2	3	4	6	5	4	4	3
IR	×	×	×	1	2	3	5	4	3	3	2
Q	×	×	×	×	1	2	4	3	2	2	3
CA	×	×	×	×	×	1	3	2	1	1	2
CO	×	×	×	×	×	×	4	3	2	2	3
DN	×	×	×	×	×	×	×	1	2	4	5
SN	×	×	×	×	×	×	×	×	1	3	4
IN	×	×	×	×	×	×	×	×	×	2	3
FC	×	×	×	×	×	×	×	×	×	×	1
GEN	×	×	×	×	×	×	×	×	×	×	×

5 Results of the function and feature analyses

We now have all the information we need to examine the results of our analyses. Table 1[6] and Table 2 represents a summary of the relations between features and functions. Table 1 shows us the feature distribution per function, and table 2 tells us how many feature values are different between every two functions. Every result is grounded on statements from Haspelmath(1997) and on applying the test from Aloni et al.(2010). Examples can be found in appendix A.

[6] SK = Specific Known, SU = Specific Unknown, IR = Irealis non-specific, Q = Question, CA = Conditional (Antecedent), CO = Comparative, DN = Direct Negation, SN = Superordinate Negation, IN = Indirect Negation, FC = Existential Free Choise and GEN = Generic.

5.1 Explaining the conceptual space

Step four in Haspelmath's typological research is explaining the induced conceptual space. The conceptual space is derived from data, but how can the arrangement of functions be explained in semantic terms? The conceptual space of Figure 9 (the same as the one in Figure 8, but pictured here again such that we have all material at hand) is, to remind us, the adapted version of Haspelmath's data-driven derived space of indefinite pronouns. With our results we can now make an effort in answering the first research question: Can the arrangement of functions in the data-driven derived conceptual space of Haspelmath(1997) be explained according to the semantic feature method?

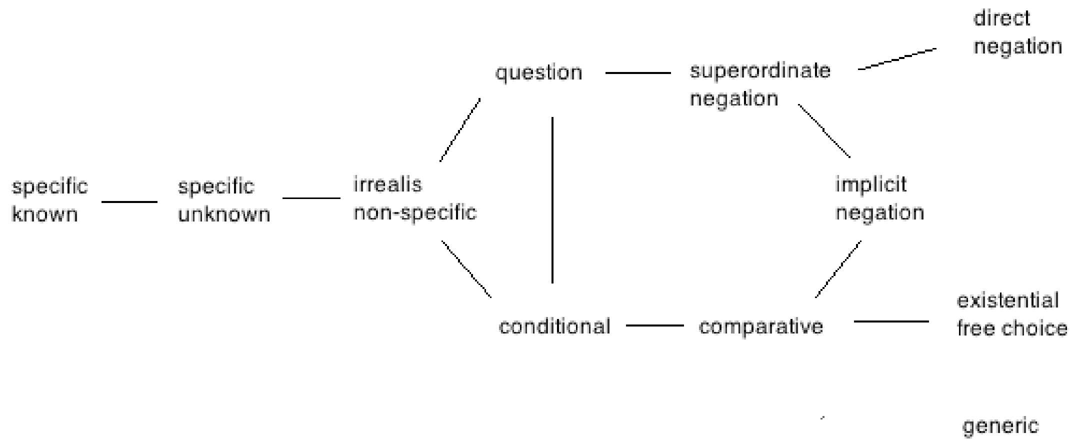

Figure 9: Conceptual space of indefinite pronouns used in this research

We can divide this research question into three subquestions : (I) Does every function correspond to a unique set of features values? (II) Do all functions that have the same value on one of the features form a contiguous area? (III) Does the number of edges between two functions correlated with the number of different feature values between them?

With the help of Table 1 and Table 2 we can conclude that the answer to (I) is positive. We found a specific set of semantic features such that every function in the conceptual space can be characterized by a unique set of feature values. This shows that the function distinction made by a data-driven approach, i.e. made by an analysis of language use, can be formulated is term of semantic features. In Figure 12-21 [7] in Appendix B it is shown that the set of functions that have the same value on one of the feature form a contiguous area, thereby answering (II) as positive as well. The positive answer to (I) and (II) are important results. It shows that there is a correspondence between data taken from language use and semantic theory based on features. There is evidently a relation between language use and the semantic properties of language. It

[7]Notice that the feature *scalar endpoint* (Figure 15) and *scale reversal* (Figure 16) is not a feature in our set

opens the wide possibilities to describe universalities and variation of language use in terms of semantic features. How it can be formulated specifically is an open question.

For (III) we wonder if the edges can also be explained in term of semantic features. Some functions are connected by one edge, while others are connected by a path of several edges. The idea is that an edge is drawn between two functions if they differ in a small number of features. Ideally, this number is fixed, but as we can conclude from Table 2 and Figure 9, this is not the case. For example, while the distance in edges between both *question* and *conditional*, and between *superordinate negation* and *question* is one, the number of different feature value is respectively one and three. However we can still explain the edges between functions in terms of feature value difference. We can use the following definition, based on a definition by Zwarts(2007: 389).

Definition 1 Two functions f_m and f_n in the conceptual space C are connected by an edge if and only if the number of different feature values between f_m and f_n is less than the number of edges of the second shortest path between f_m and f_n.

This definition states that the number of edges of the shortest path between two function is less than or equal to the number of different feature values between those functions. For example, the edge between *superordinate negarion* and *question* (feature value difference is three) is consistent with the definition, because the second shortest path between them is more than three edges. This statement is satisfied for every connection in the conceptual space, except two. The feature difference between *conditional* and *existential free choice*, and between *question* and *existential free choice* is less than the number of edges between them. The solution would be an edge between *conditional* and *existential free choice*. This means that with our set of features we can not completely formulate the edges in the conceptual space for indefinite pronouns, thereby giving a negative answer to (III).

So far we formulated a nearly complete explanation for differentiation and arrangement of the the functions in the conceptual space based on the semantic featured method, thereby providing the right path for a (largely) positive answer to our first research question. However, there are still some other open issues. An interesting question is whether all contiguous areas can be defined in terms of a set of feature values. For instance the contiguous area of *direct negation*, *superordinate negation* and *implicit negation* can exclusively defined by feature value N+. Other areas can be defined by a combination or set of feature values. The set {N–, A+, R+} divides the area of *question*, *comparative* and *conditional* from the rest. Unfortunately, it is not the case that every contiguous area can be characterized by a set of feature values. We can provide at least one counterexample. Namely, the area with the functions *question*, *superordinate negation* and *implicit negation* [8]. Another interesting question is whether only contiguous area's can

[8]The whole set of features the three functions have in common is S–, K–, A+, D-, C-, E+ and R+. But this set is also includes *conditional*, which means there is no characteristic set for the three functions.

be characterized by a set of feature values. Again, this is not the case. A counterexample is the non-contiguous area of *Conditional*, *Existential Free Choice* and *Implicit Negation*, that can be characterized by the set { ∀+, A+, M-, C-}.

Our feature based analysis mostly cover the function arrangement of Haspelmath's conceptual spaces. The exceptions show that there are still some unsolved issues in the correspondence between the data and the semantic theory explaining the data. We will see that the last two cases, and the fact that the edges aren't fully explained by semantic features, disappear when we derived a conceptual space from a semantic analysis. This is outlined in the next section.

5.2 Function arrangement based on function and feature analysis

We made an comprehensive feature based semantic analysis of every function in the semantic map of indefinite pronouns. Now, let's use this analysis to deduces an alternative conceptual space, without looking at the data. This conceptual space is shown in Figure 10. Every edge between two functions corresponds with a difference of only one feature value between that two functions. The double edge between *irrealis non-specific* and *generic* is officially not part of the semantic map method, but is an ad-hod way of keeping up the correspondence between number of edges and number of different feature values (as seen in Table 2) between functions . Figure 10 is completely consistent with Table 2, and with Definition 1.

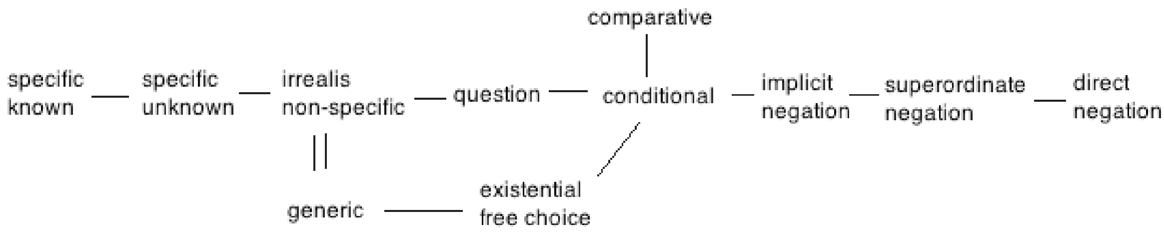

Figure 10: Conceptual space deduced from a semantic analysis of features

This brings us to the second research question: when a conceptual space is deduced from a space-driven perspective using the feature method, what are the similarities and differences with the empirically induced conceptual space of Haspelmath? When comparing the conceptual space of Figure 9 with the one of Figure 10, some fundamental differences can be observed. Figure 11 pictures them both next to each other. The two are still roughly the same: *specific known (SK)*, *specific unknown (SU)* and *irrealis non-specific (IR)* form a contiguous area; this is also the case for *question (Q)*, *comparative (CO)* and *conditional (CA)*, for *indirect negation (IN)*, *superordinate negation (SN)* and *direct negation (DN)*, and for *existential free choice (FC)* and *generic (GE)*. The most important difference between the two conceptual space, is that some set of functions that form a contiguous area in Figure 11a (the adapted version of Haspelmath), don't form a contiguous area in FIgure 10b (our semantically derived version). The same is the case for the other way around. It is this fact that explain the problem mentioned at the end of section 5.1. The contiguous area *question*, *superordinate negation* and *implicit negation* in Haspelmath's space can't be defined by a set of feature values, probably because they don't form a contiguous areas in the conceptual space that is derived from those features. In the same way that the non-contugious area *conditional*, *existential Free Choice* and *implicit Negation* in Haspelmath's map probably can be defined by a set of feature values, because they form a contiguous area in the semantically derived map. However, we still need an official proof that all and only contiguous areas in Figure 11b can be defined by a set of feature value, that differentiates it from the rest, but we will leave that for future research.

Overall, the conceptual space derived from data and the space derived from a semantic analysis are very similar. Although it must be said the two approaches in this thesis were not totally separated from each other. For efficiency purposes we didn't semantically analyze the indefinite pronouns, but rather the functions óf indefinite pronouns. The functions that were already derived from a data driven perspective. Also our data-driven conceptual space had one function distinction based on semantic ground; *superordinate negation* and *implicit negation.* However, these minor issues cannot deny the fact that there is a strong correspondence between the way language is used and an explanation of that use in semantical terms.

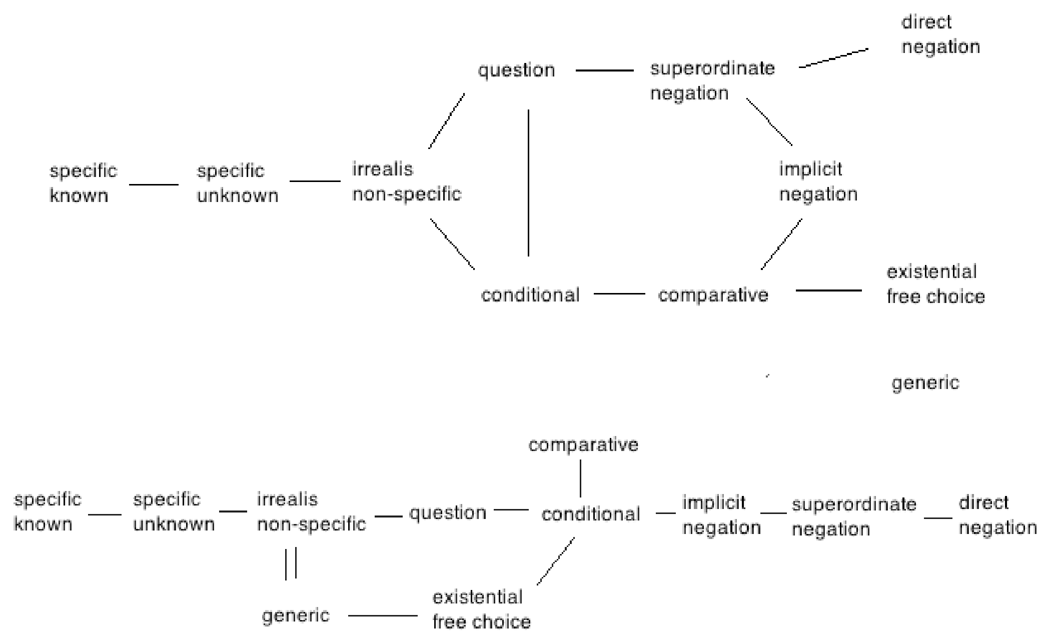

Figure 11: (a): conceptual space induced from data (above) and (b) conceptual space deduced from a semantic analysis of features (beneath)

6 Conclusion

Haspelmath's work on semantic maps is seen as a important piece of research within the field of language typology. His book about indefinite pronouns was the starting point and main source of information for this research. The four steps of typological research described by Haspelmath, and the two different approaches for deriving a conceptual space described by Zwarts(2007) formed the main structure for this thesis.

Haspelmath used a purely data-driven perspective for deriving his conceptual space for indefinite pronouns. The selection and arrangement of functions in the space were based on data; it was a representation on how language is used. Haspelmath's semantical explanation based on features wasn't satisfiable and needed some further investigation. In this research we found a set of features that extended Haspelmath's explanation and covered nearly all data. We used the same set of features to derive a conceptual space from another perspective. Analyzing the functions of indefinite pronouns semantically lead to a slightly different space.

The data-driven semantic map shows the universal and variational properties of how language is used. The given semantic explanation and the similarities between the two derived conceptual spaces open the possibilities to formulate the characteristics of language use in terms of features. Thereby, showing a strong correspondence between language use (the data) and the semantical (and syntactical, pragmatical and logical) properties of language for indefinite pronouns. This is an important result that is compatible with other fields of research within cognitive science and artificial intelligence. For instance formal semantics, where language use is modeled based on logical properties. Or machine learning and automated text processing, where texts must be annotated, i.e linguistic properties are assigned to words. To conclude, the result of this thesis confirms the strong correspondence between language use and semantic, syntactic, pragmatic and logical properties of language.

7 References

Aloni, Maria, Ana Aguilar Guevara, Angelika Port, Radek Šimík, Machteld de Vos and Hedde Zeijlstra. 2010. Indefinites as fossils: a synchronic and diachronic appraoch corpus study.

Cysouw, Michael. 2007. Building semantic maps: The case of person marking. New Challenges in typology, ed. by Matti Miestamo and Bernhard Wälchli, 225-248. Berlin: Mouton.

Goddard, Cliff. 2010. The Natural Semantic Metalanguage Approach. In: Bernd Heine and Heiko Narrog (eds.), The Oxford Handbook of Linguistic Analysis. Oxford: Oxford University Press: 459-484. De Haan, Ferdinand. 2004. On representing semantic maps.

Haspelmath, Martin. Indefinite pronouns. Oxford University Press, Oxford 1997.

Haspelmath, Martin. 2003. The Geometry of Grammatical Meaning: Semantic Maps and Cross-Linguistic Comparison. The new psychology of language, ed. by Michael Tomasello, vol 2, 211-243. New York: Erlbaum.

Van der Wouden, Ton. 1994. Negative Context. Groningen Dissertations in Linguistics.

Zwarts, Joost. 2007. Semantic Map Geometry: Two approaches. Linguistic Discovery 8.1: 377-395

8 Appendix A

zin	features
1) Nobody came.	S–, K–, ∀+, Q–, A+ G-, N+, M+, D+, C–
2) I noticed nothing.	S–, K–, ∀+, Q–, A+, G-, N+, M+, D+, C–
3) The cowboy was nowhere to be found.	S–, K–, ∀+, Q–, A+ G-, N+, M+, D+, C–
4) I can't find it anywhere.	S–, K–, ∀+, Q–, A+, G-, N+, M+, D+, C–
5) I have not seen any of them.	S–, K–, ∀+, Q–, A+, G-, N+, M+, D+, C–
6) I will not say anything.	S–, K–, ∀+, Q–, A+, G-, N+, M+, D+, C–
7) I don't think that anybody come(s?).	S–, K–, ∀+, Q–, A+, G-, N+, M+, D–, C–
8) There isn't a reason why anyone would buy that	S–, K–, ∀+, Q–, A+ G-, N+, M+, D–, C–
9) I don't believe that anybody has seen it.	S–, K–, ∀+, Q–, A+, G-, N+, M+, D–, C–
10) The company refuses to make any decision.	S–, K–, ∀+, Q–, A+, G-, N+, M–, D–, C–
11) The security will deny access to anybody.	S–, K–, ∀+, Q–, A+ G-, N+, M–, D–, C–

zin	features
1) He didn't do anything to help her	S–, K–, ∀+, Q–, A+, G–, N+, M+, D+, C–
2) I didn't see anyone	S–, K–, ∀+, Q–, A+, G–, N+, M+, D+, C–
3) If you tell anybody, we'll punish you	S–, K–, ∀+, Q–, A+, G–, N–, M–, D–, C–
4) If you see anybody, tell me immediately	S–, K–, ∀+, Q–, A+, G–, N–, M–, D–, C–
5. If you go anywhere, send me a card	S–, K–, ∀+, Q–, A+, G–, N-, M–, D–, C–
6) Few people show any interest in global issues.	S–, K–, ∀+, Q–, A+ G–, N+, M–, D–, C–
7) Only Mira has anything substantial to report.	S–, K–, ∀+, Q–, A+, G–, N+, M–, D–, C–
8) Khadija is too occupied to invite anyone else	S– , K–, ∀+, Q–, A+, G–, N+, M–, D–, C–
9) We doubt that Juan applied anywhere	S–, K–, ∀+, Q–, A+, G–, N+, M–, D?, C–
10) Without saying anything, she ate the soup	S–, K–, ∀+, Q–, A+, G–, N+, M–, D–, C–
11) The boy can run faster than anyone in his class.	S–, K–, ∀+, Q–, A+, G–, N–, M–, D–, C+
12) The boy runs as fast as anyone in his class	S–, K–, ∀+, Q–, A+, G–, N–, M–, D–, C+
13) This place is better than anywhere else	S–, K–, ∀+, Q–, A+, G–, N–, M–, D–, C+
14) Everyone who likes any kind of seafood will like our new seafood sticks	S–, K–, ∀+, Q–, A+, G–, N–, M–, D–, C–
15) Is there anything new?	S–, K–, ∀–, Q+, A+, G–, N–, M–, D–, C–
16) Did you see anything?–	S–, K–, ∀–, Q+, A+, G–, N–, M–, D–, C–

zin	features
1) Ivan wants to sing some romance	S+-, K+-, ∀-, Q-, A-, G–, N-, M–, D–, C-
2) She wanted to acquire some object	S+-, K+-, ∀-, Q-, A-, G–, N-, M–, D–, C-
3) He wants to marry some Greek woman	S+-, K+-, ∀-, Q-, A-, G–, N-, M–, D–, C-
4) Somebody came	S+, K+-, ∀-, Q-, A-, G– , N-, M–, D–, C-
5) Look, somebody is running	S+, K+-, ∀-, Q-, A- G–, N-, M, D–, C-
6) On Saturday they will go somewhere^{specific}	S+, K+-, ∀-, Q-, A-, G–, N-, M–, D–, C-
7) Someone^{specific} can come	S+, K+-, ∀-, Q-, A- G–, N-, M–, D–, C-
8) I will tell you if anyone comes	S-, K-, ∀+, Q-, A+ G+, N-, M–, D–, C-
9) Did you see anyone?	S-, K-, ∀-, Q+, A+ G+, N-, M–, D–, C-
10) Visit me sometime	S-, K-, ∀-, Q-, A- G–, N-, M–, D–, C
11) Go and find some sailor	S-, K-, ∀-, Q-, A- G–, N- , M–, D–, C-
12) Apperently someone complained about her	S-, K-, ∀-, Q-, A- G–, N-, M–, D–, C-
13) On Saturday they will go somewhere$^{non-specific}$	S-, K-, ∀-, Q-, A- G– ,N- , M–, D–, C-
14) Someone$^{non-specific}$ can come	S-, K-, ∀-, Q-, A- G–, N-, M–, D–, C-

zin	features
1) I can write articles anywhere	S-, K-, ∀+, Q-, A+, G–, N-, M–, D–, C-
2) any pianist can play a piece like that	S-, K-, ∀+, Q-, A+ G+, N-, M–, D–, C-
3) you can take any apple	S-, K-, ∀+, Q-, A+, G+, N-, M–, D–, C-
5) Go ask anyone. You'll see that I'm right	S-, K-, ∀+, Q, A+, G+, N-, M–, D–, C-
6) Ask anybody	S-, K-, ∀-, Q-, A- G+, N-, M–, D–, C-
7) Petja will solve any problem	S-, K-, ∀+, Q-, A+ G+, N-, M–, D–, C-
8) Any cat is an animal	S-, K-, ∀+, Q-, A+, G+, N-, M–, D–, C-
9) I would give anything to see that	S-, K-, ∀+, Q-, G+, N-, M–, D, C-
10) She would talk to anyone about her love problems	S-, K-, ∀+, Q-, G–, N-, M–, D, C-
11) Any amount is sufficient	S-, K-, ∀+, Q-, G–, N-, M–, D, C-

9 Appendix B

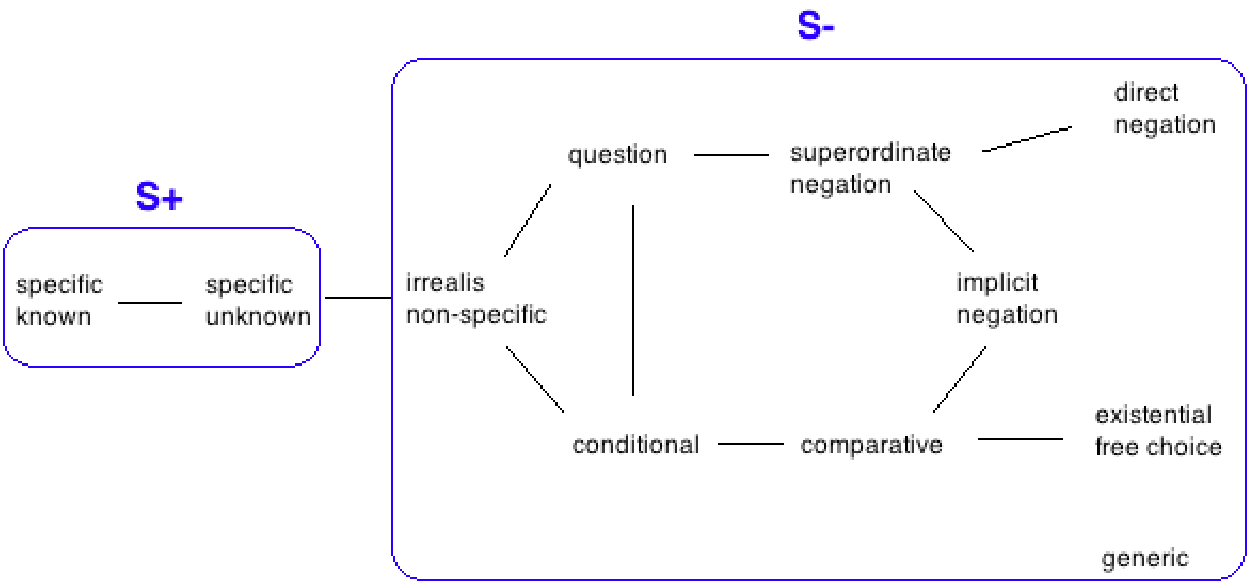

Figure 12: distribution of specifity over the conceptual space of indefinite pronouns

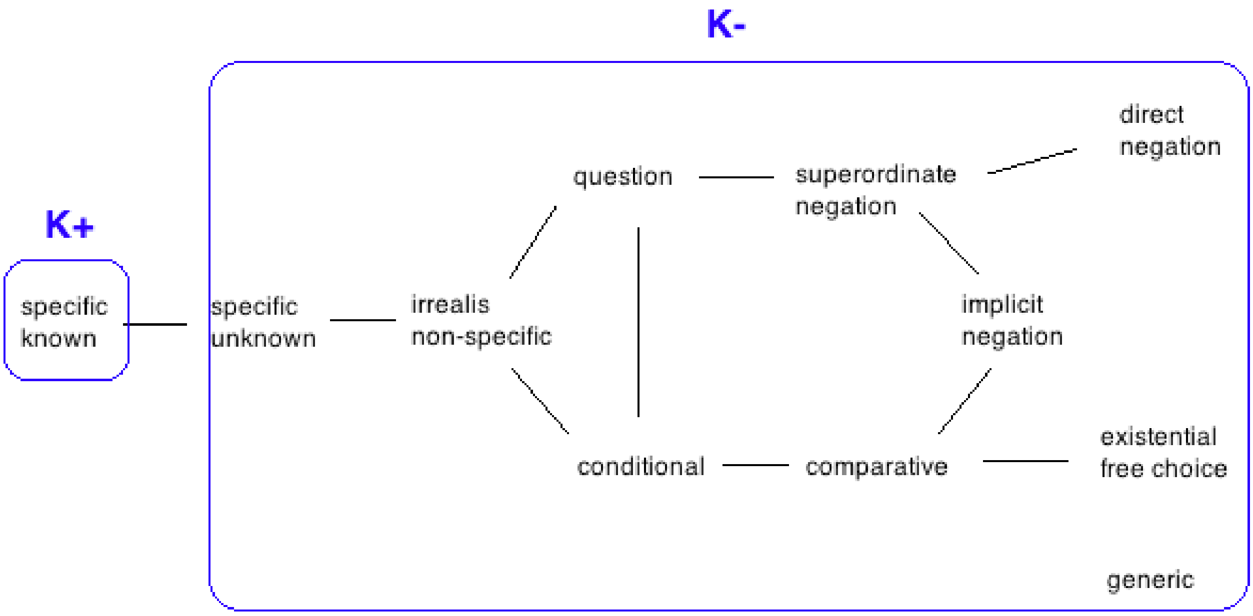

Figure 13: distribution of known over the conceptual space of indefinite pronouns

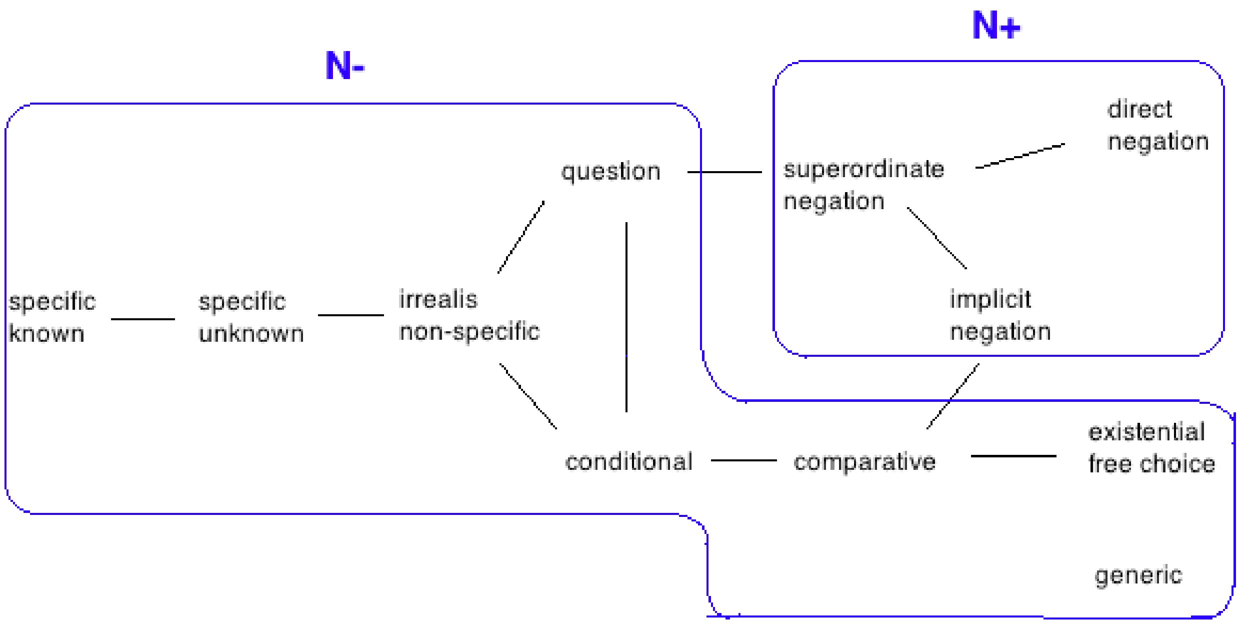

Figure 14: distribution of negative meaning over the conceptual space of indefinite pronouns

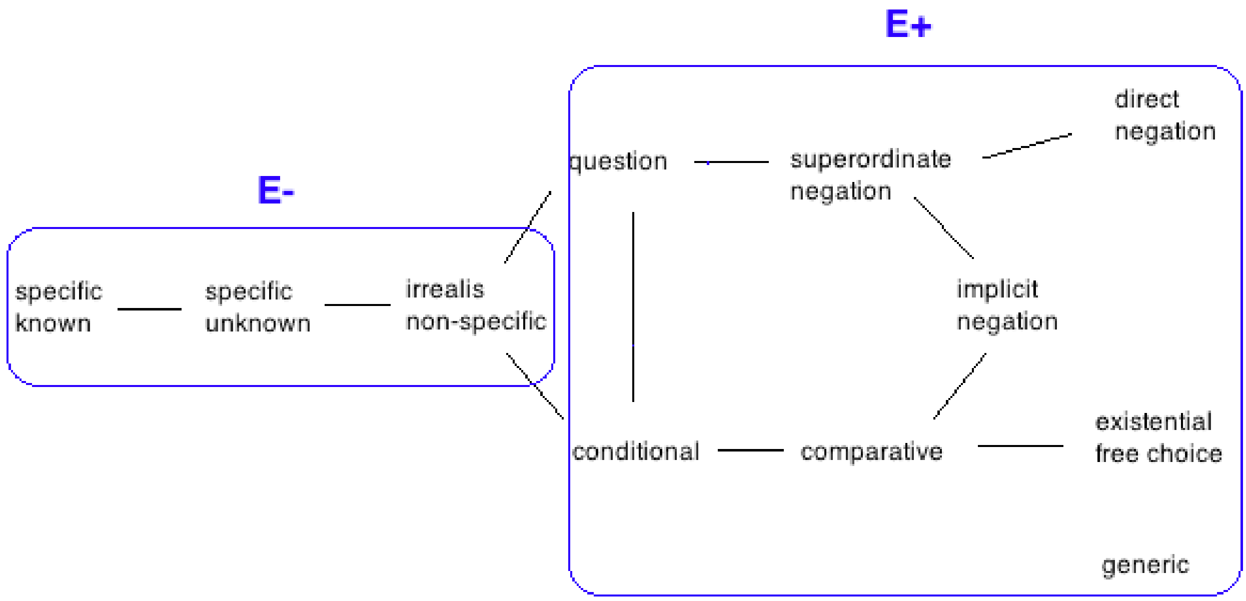

Figure 15: distribution of scalar endpoint over the conceptual space of indefinite pronouns

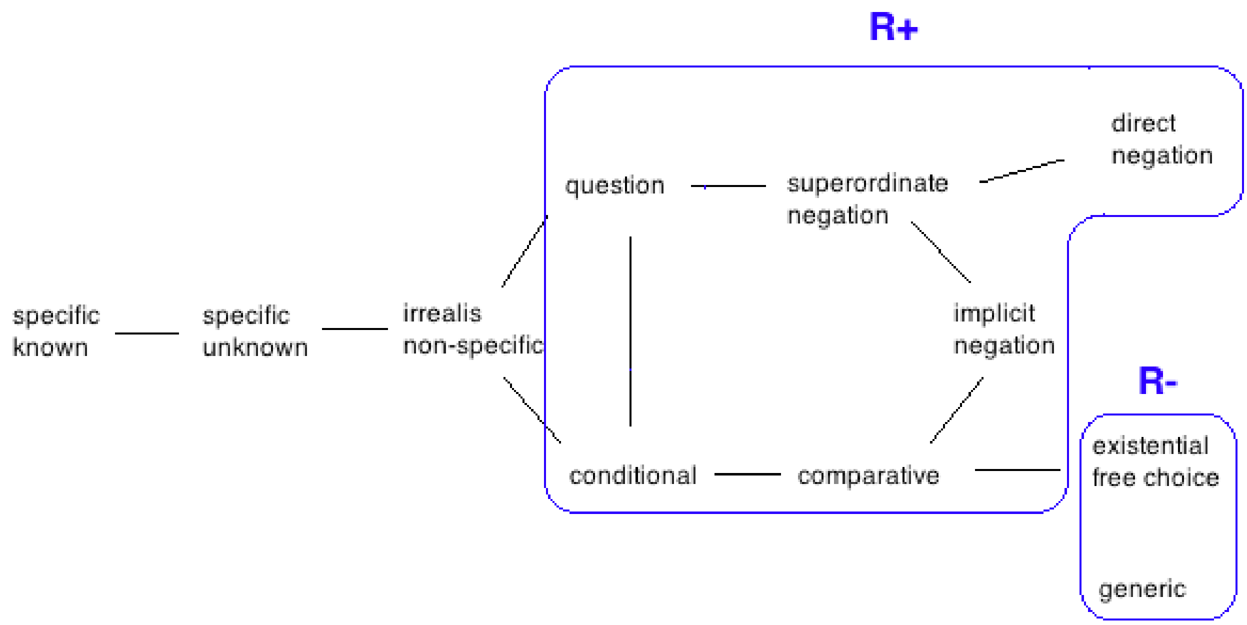

Figure 16: distribution of scale reversal over the conceptual space of indefinite pronouns

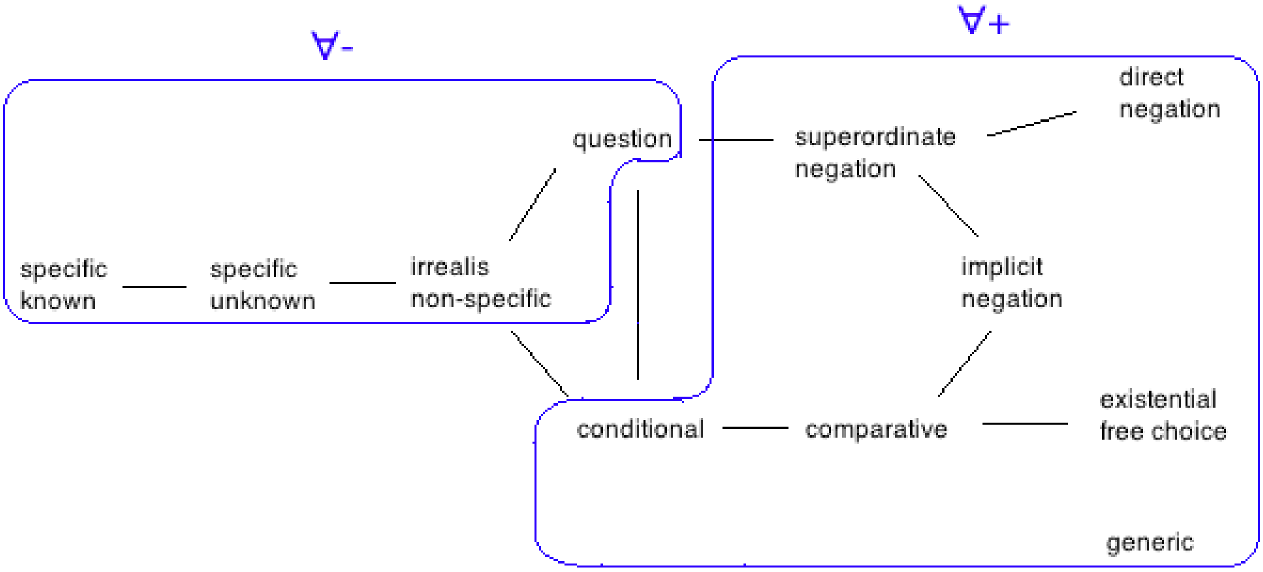

Figure 17: distribution of universal meaning over the conceptual space of indefinite pronouns

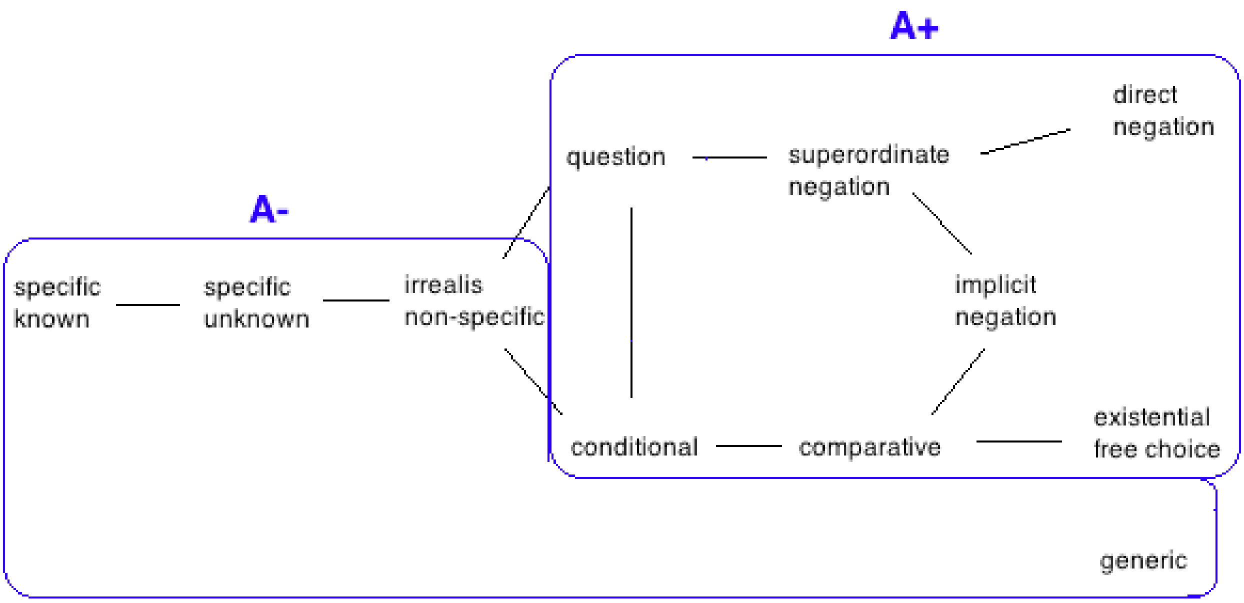

Figure 18: distribution of anti-additivity over the conceptual space of indefinite pronouns

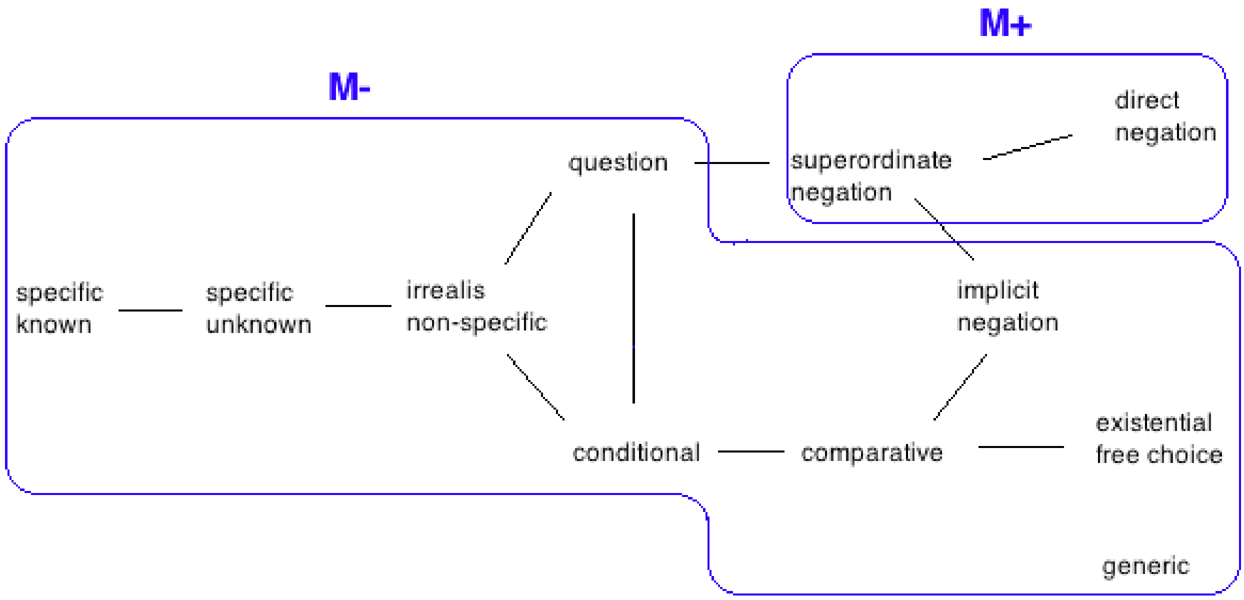

Figure 19: distribution of anti-multiplicativity over the conceptual space of indefinite pronouns

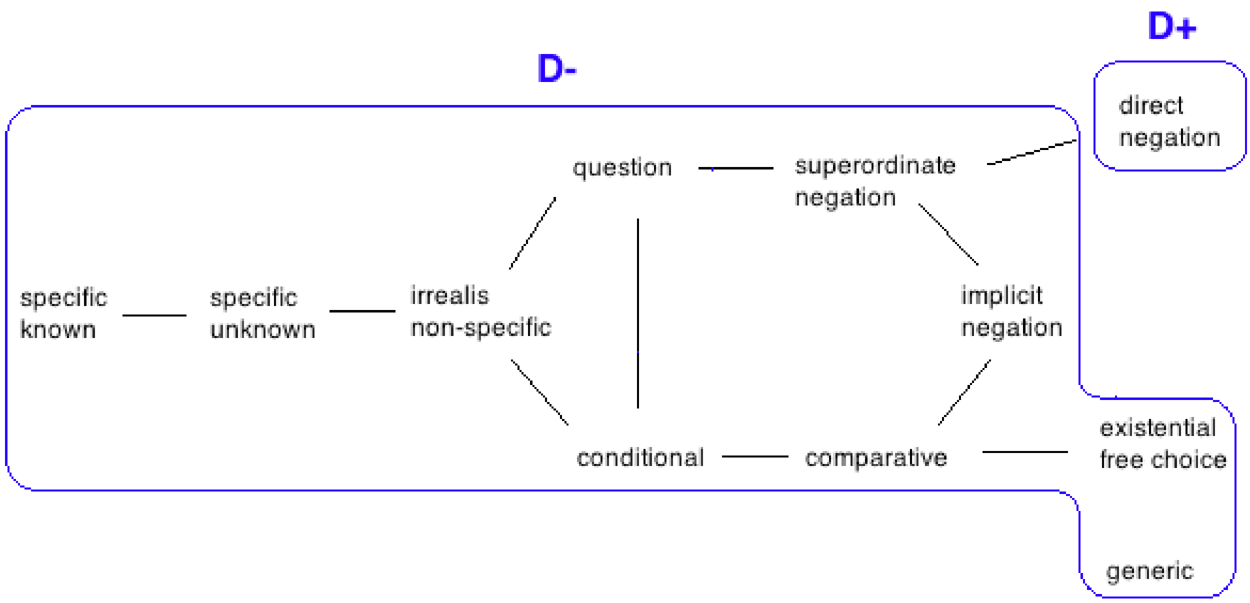

Figure 20: distribution of clausal negation over the conceptual space of indefinite pronouns

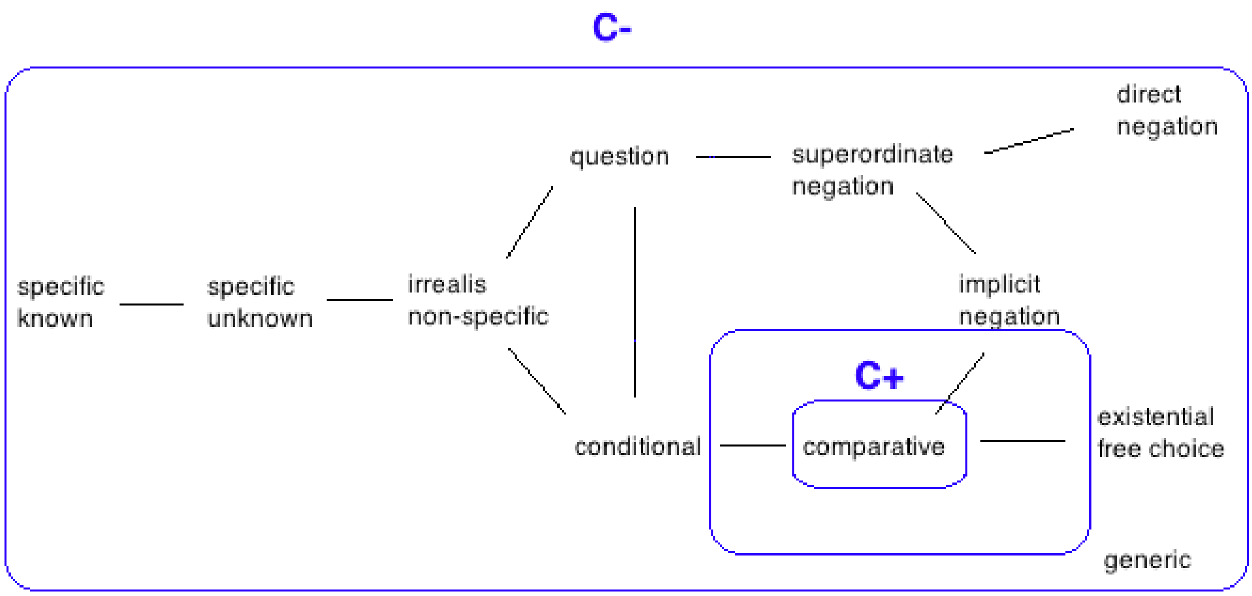

Figure 21: distribution of comparative over the conceptual space of indefinite pronouns