

Semantics and Pragmatics of Indefinites: Methodology for a Synchronic and Diachronic Corpus Study

Ana Aguilar-Guevara^{*}, Maria Aloni[†], Angelika Port[†], Radek Šimík[‡],
Machteld de Vos^{**}, Hedde Zeijlstra[†]

^{*}Utrecht University, [†]University of Amsterdam, [‡]University of Potsdam,
^{**}Cambridge University

Abstract

The article discusses the methodology adopted for a cross-linguistic synchronic and diachronic corpus study on indefinites. The study covered five indefinite expressions, each in a different language. The main goal of the study was to verify the distribution of these indefinites synchronically and to attest their historical development. The methodology we used is a form of functional labeling which combines both context (syntax) and meaning (semantics) using as a starting point Haspelmath's (1997) functional map. In the article we identify Haspelmath's functions with logico-semantic interpretations and propose a binary branching decision tree assigning each instance of an indefinite exactly one function in the map.

1 Theoretical Background

It is well known that the use of expressions with existential meaning (e.g. plain indefinites like English *somebody*, or Czech *někdo*) can give rise to different pragmatic effects. Although the semantic representation of *somebody* in (1) and (2) is identical, (1) comes along with a **free choice implicature** (each individual is a permissible option) and (2) with an **ignorance implicature** (the speaker does not know who called):

- (1) You can invite somebody.
- (2) Somebody called.

From a typological perspective, many languages have developed specialized forms for such enriched meanings, such as **free choice indefinites**¹: Spanish *cualquier*-series, Czech *koli*-series, Dutch *dan ook*-series, . . . , and as **epistemic indefinites**²: Russian *to*-series, Czech *si*-series, German *irgend*-series, Spanish *algun*-series, . . .

Following Grice's seminal work, the main hypothesis that motivates the present research is that these different indefinite forms have emerged as result of a process of conventionalization (or fossilization) of an originally pragmatic inference.

In languages with Epistemic Indefinite (EI) forms, inference (3c), pragmatic in origin, has been integrated into the semantic content of sentences like (4a).

¹E.g. Dayal (1998), Giannakidou (2001), Menéndez-Benito (2010).

²E.g. Kratzer and Shimoyama (2002), Jayez and Tovena (2006), Alonso-Ovalle and Menéndez-Benito (2010).

(3) *Plain indefinite (German)*

- a. **Jemand** hat angerufen.
somebody has called
- b. Conventional meaning: Somebody called
- c. Ignorance implicature: The speaker does not know who

(4) *EI pronoun (German 'irgendjemand')*

- a. **Irgendjemand** hat angerufen.
somebody:UNKNOWN has called
- b. Conventional meaning: Somebody called and the speaker does not know who

In languages with distinctive Free Choice (FC) forms, inference (5c) pragmatic in origin, has been integrated into the semantic content of sentences like (6a).

(5) *Plain indefinite (Spanish)*

- a. *Puedes traer un libro.*
can:2SG bring:INF a book
- b. Conventional meaning: You can bring me a book
- c. Free choice implicature: Each book is a possible option

(6) *FC determiner (Spanish 'cualquier')*

- a. *Puedes traer cualquier libro.*
can:2SG bring:INF any book
- b. Conventional meaning: You can bring me a book and each book is a possible option

In this project, cross-linguistic synchronic and diachronic studies have been combined in order to substantiate this hypothesis. The synchronic studies intend to determine what has been fossilized, the diachronic studies how this has happened.

In the synchronic research we studied the following indefinite forms: German EI *irgendein*, Czech FC *kterýkoli*, Italian FC *(uno) qualunque*, Spanish FC *cualquiera* and Dutch FC *wie dan ook*. The main goal of this research was to understand which part of the meaning of the indefinite form is fossilized and to develop some hypotheses on how it might have happened diachronically. In the diachronic corpus research we studied the historical development of the last two indefinite forms: Spanish *cualquiera* and Dutch *wie dan ook*.

In this article we will focus on the methodology developed for these corpus studies, and report on parts of the diachronic research as an illustration of our results.

2 Corpus study: diagnostics and methodology

In the synchronic and diachronic studies we have classified randomly selected occurrences of each indefinite according to a number of categories. The annotation was carried out by five annotators (one per language) who met regularly to compare their results and share their experience with the annotation instructions.³ The starting point

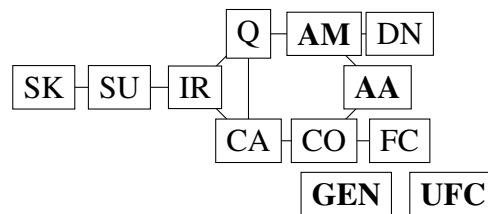
³An assessment of the methodology by measuring inter-annotator agreement with the *kappa* coefficient has been carried out in January 2011. Five annotators coded 100 randomly chosen examples from the British National Corpus. Each example contained one marked occurrence of *some* (20 examples) or *any* (80 examples). The average kappa score obtained was 0.52, with a standard deviation of 0.069. We performed a second calculation where the disagreements among the three negative labels (AA, AM and DN) and among the two specific labels (SK and SU) were not taken into account (had a weight of 0), and where the disagreements between the specific functions and IR were considered half correct (had a weight of 0.5). This yielded a kappa score of 0.69, with a standard deviation of 0.106 (for details see van Cranenburgh et al. 2011).

for the identification of the relevant categories was Haspelmath’s functional map. In this section, we introduce our extended version of Haspelmath’s map and provide an explicit set of logico-semantic criteria, according to which indefinites are assigned functions on the map.

2.1 Haspelmath’s semantic map

Haspelmath’s (1997) typological survey identified 9 main functions for indefinite forms organized in an implicational map. We will assume the following extended version of Haspelmath’s map motivated by a more detailed NPI/FC classification (Aguilar-Guevara et al. 2010). The newly introduced functions are in boldface in the following illustrations:

(7) *An extended version of Haspelmath’s map*



(8) *Functions on the map*

	Abbr	Label	Example
a.	SK	specific known	<i>Somebody</i> called. Guess who?
b.	SU	specific unknown	I heard <i>something</i> , but I couldn’t tell what it was.
c.	IR	irrealis	You must try <i>somewhere</i> else.
d.	Q	question	Did <i>anybody</i> tell you anything about it?
e.	CA	conditional antecedent	If you see <i>anybody</i> , tell me immediately.
f.	CO	comparative	John is taller than <i>anybody</i> .
g.	DN	direct negation	John didn’t see <i>anybody</i> .
h.	AM	anti-morphic	I don’t think that <i>anybody</i> knows the answer.
i.	AA	anti-additive	The bank avoided taking <i>any</i> decision.
j.	FC	free choice	You may kiss <i>anybody</i> .
k.	UFC	universal free choice	John kissed <i>any</i> woman with red hair.
l.	GEN	generic	<i>Any</i> dog has four legs.

In order for an indefinite to qualify for a function, it must (i) be grammatical in the context the function specifies; and (ii) have the semantics that the function specifies. For example, *any* does not exhibit the specific functions SK/SU because it is ungrammatical in episodic sentences, cf. (9a); and *some* does not exhibit the comparative function CO because it does not have a universal meaning specified by CO, cf. (9b).

- (9) a. He went somewhere /# anywhere else.
 b. Berlin is bigger than any /# some Czech city.
 ‘For all Czech cities it holds that Berlin is bigger than they are.’

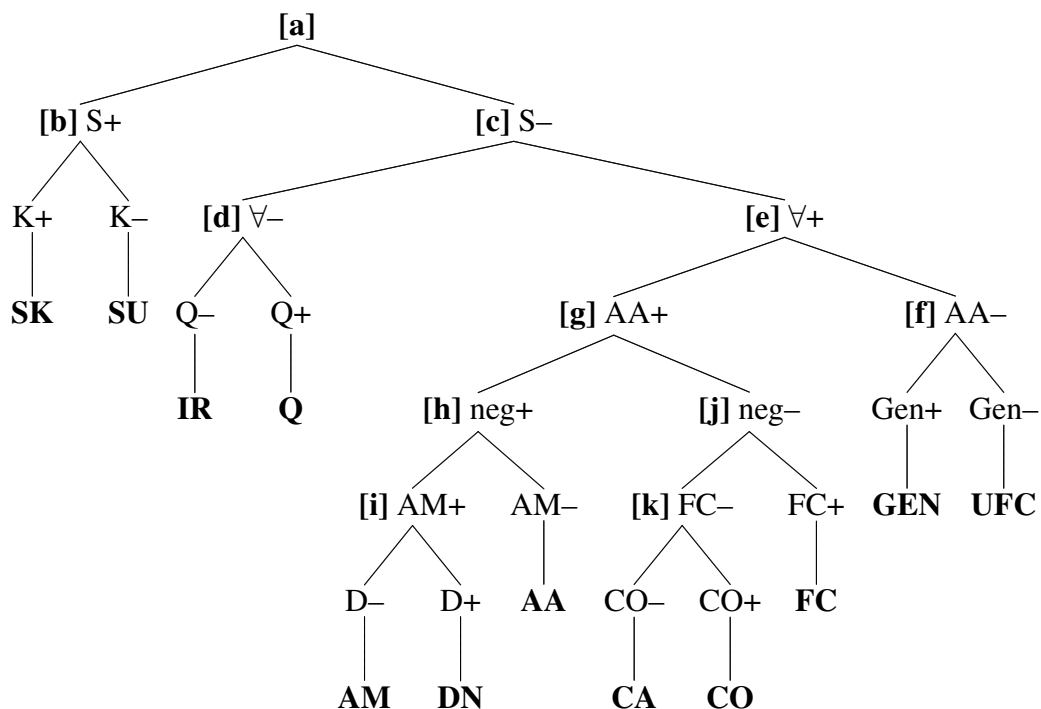
Epistemic indefinites are indefinites that exhibit the SU function, but not the SK function. Free choice indefinites are indefinites exhibiting the FC function.

Haspelmath proposes that an indefinite will always express a set of functions that are contiguous on the map (where two functions are contiguous iff they are connected by a line).⁴ One prediction is that items which acquire new functions will develop first those functions that are contiguous to the original function.

2.2 Methodology for semantic annotation

In this section we introduce a set of tests which we used to assign exactly one function to each instance of the examined indefinites. These tests and the order in which they were applied are schematized in the following decision tree.

(10) *Decision tree*



For each node in the decision tree we give now the corresponding test, and, as an illustration, we apply it to the sentences we have used in (8) to exemplify our functional labels. Our first test is test (a) used to distinguish specific from non-specific uses of indefinites.

(a) Test for specificity [S+/-]:

Sentence (S): ...indefinite_i ... **Possible Continuation (PC):** ... pronoun_i ... [S+]

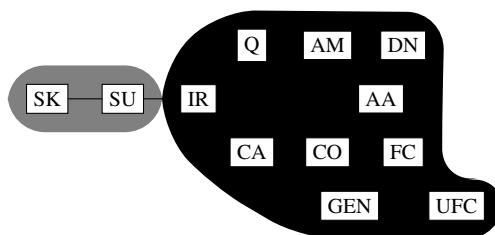
Examples:

- | | |
|---|------|
| a. <i>Somebody_i</i> called. She _i wanted a new appointment. | [S+] |
| b. I heard <i>something_i</i> . It _i was very loud. | [S+] |
| c. You must try <i>somewhere_i</i> else. # It _i is a very nice place. | [S-] |
| d. Did <i>anybody_i</i> tell you anything about it? # He _i is a real chatterbox. | [S-] |
| e. If you see <i>anybody_i</i> , tell me immediately. # He _i is a nice guy. | [S-] |

⁴The precise placement on the map (i.e. connecting lines determining function contiguity) of GEN and UFC is still a matter of investigation.

- f. John is taller than *anybody*_i. # He_i is short. [S-]
 g. John didn't see *anybody*_i. # He_i was very tall. [S-]
 h. I don't think that *anybody*_i knows the answer. # He_i did not even try. [S-]
 i. The bank avoided taking *any* decision_i. # It_i was difficult. [S-]
 j. You may kiss *anybody*_i. # She_i is beautiful. [S-]
 k. John kissed *any* woman_i with red hair. # She_i is Italian. [S-]
 l. *Any* dog_i has four legs. # It_i is very cute. [S-]

The application of test (a) splits our map into a specific area (in grey) and a non-specific area (in black).



Within the specific area we apply test (b) to distinguish the specific known from the specific unknown function.

(b) Test for known [K+/-]: S: ... indefinite ... PC: Guess who/what? [K+]

Examples:

- a. *Somebody* called. Guess who? [K+] \mapsto [SK]
 b. I heard *something*, but I couldn't tell what it was. # Guess what? [K-] \mapsto [SU]

Within the non-specific area we apply test (c) to distinguish between wide-scope universal meaning and genuinely existential meaning:

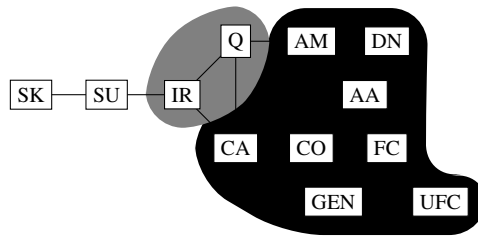
(c) Test for universal meaning [\forall +/-]:

... **Op** (... indefinite ...) ... \Rightarrow ... $\forall x$ (**Op**... x ...) ...

Examples:

- a. You must try *somewhere* else \nRightarrow for every place x : you must try x [\forall -]
 b. Did *anybody* tell you anything about it? \nRightarrow for every x : did x tell you about it? [\forall -]
 c. If you see *anybody*, tell me immediately \Rightarrow for every x : if you see x , tell me immed. [\forall +]
 d. John is taller than *anybody* \Rightarrow for every x : John is taller than x [\forall +]
 e. I didn't see *anybody* \Rightarrow for every x : I didn't see x [\forall +]
 f. I don't think that *anybody* knows the answer \Rightarrow for every x : I don't think that x knows the answer [\forall +]
 g. The bank avoided taking *any* decision \Rightarrow for every decision x : the bank avoided taking x [\forall +]
 h. You may kiss *anybody* \Rightarrow for every x : you may kiss x [\forall +]
 i. John kissed *any* woman with red hair \Rightarrow for every woman x with red hair: John kissed x [\forall +]
 j. *Any* dog has four legs \Rightarrow for every dog x (with exceptions?): x has four legs [\forall +]

The application of test (c) splits the non-specific area into an existential area (in grey) and a wide-scope universal area (in black).



Within the existential area we distinguish polar questions from irrealis non-specific constructions via step (d).

(d) Polar question [Q+]

Examples:

- a. You must try *somewhere* else. [Q-] ↦ [IR]
- b. Did you see *anybody*? [Q+] ↦ [Q]

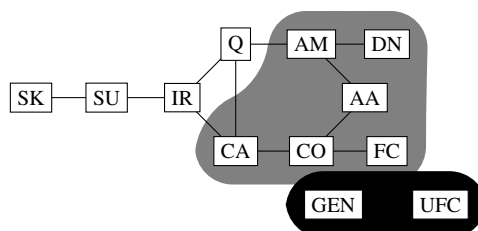
Within the wide-scope universal area we apply test (e) to distinguish anti-additive contexts from non anti-additive ones.

(e) Test for anti-additivity [AA+/-]: $\text{Op}(a \vee b) \Rightarrow \text{Op}(a) \wedge \text{Op}(b)$ [AA+]

Examples:

- a. If you see *anybody*, you should tell me immediately. [If you see John or Mary, you should tell me immediately \Rightarrow If you see John, you should tell me immediately and if you see Maria, you should tell me immediately] [AA+]
- b. John is taller than *anybody*. [John is taller than Lee or Mary \Rightarrow John is taller than Lee and John is taller than Mary] [AA+]
- c. John didn't see *anybody*. [John didn't see Lee or Mary \Rightarrow John didn't see Lee and John didn't see Mary] [AA+]
- d. I don't think that *anybody* knows the answer. [I don't think that Mary or Lee know the answer \Rightarrow I don't think that Mary knows the answer and I don't think that Lee knows the answer] [AA+]
- e. The bank avoided taking *any* decision. [The bank avoided taking decision A or decision B \Rightarrow The bank avoided taking decision A and the bank avoided taking decision B] [AA+]
- f. You may kiss *anybody*. [You may kiss John or Mary \Rightarrow you may kiss John and you may kiss Mary] [AA+]
- g. John kissed *any* woman with red hair. [John kissed Lee or Bea $\not\Rightarrow$ John kissed Lee and John kissed Bea] [AA-]
- h. *Any* dog has four legs. [Fido or Bobby has four legs $\not\Rightarrow$ Fido has four legs and Bobby has four legs] [AA-]

The application of test (e) splits the universal area into an anti-additive area (in grey) and a non anti-additive area (in black).



Within the non anti-additive area we apply test (f) to distinguish generic from universal free choice readings.

(f) Test for genericity [Gen+/-]: ...indefinite ... ≡ ... plain generic indef. ... [Gen+]

Examples:

- a. John kissed *any* woman with red hair \neq John kissed a woman with red hair
[Gen-] \mapsto [UFC]
- b. *Any* dog has four legs \equiv A dog has four legs
[Gen+] \mapsto [GEN]

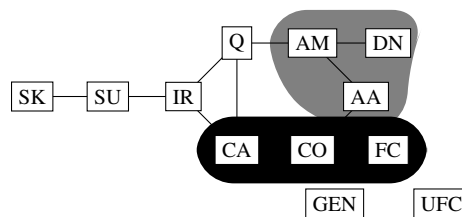
Within the anti-additive area we apply test (g) to distinguish negative contexts from non negative ones.

(g) Test for negative meaning [Neg+/-]: $\text{Op}(a \vee \neg a)$ is inconsistent [Neg+]

Examples:

- a. John didn't see *anybody*. [John didn't stay or go \mapsto inconsistent] [Neg+]
- b. I don't think that *anybody* knows the answer. [I don't think that the door is open or closed \mapsto inconsistent] [Neg+]
- c. The bank avoided taking *any* decision. [The bank avoided being open or closed] \mapsto inconsistent] [Neg+]
- d. You may kiss *anybody*. [You may stay or go \mapsto not inconsistent] [Neg-]
- e. If you see *anybody*, you should tell me. [If you stay or go, you should tell me \mapsto not inconsistent] [Neg-]
- f. John is taller than *anybody*. [John is taller than somebody or nobody \mapsto not inconsistent] [Neg-]

The application of test (g) splits the anti-additive area into a negative area (in grey) and a non-negative area (in black).



Within the negative area we apply test (h) to distinguish anti-multiplicative contexts from plain negative ones.

(h) Test for anti-multiplicativity: $\text{Op}(a) \vee \text{Op}(b) \equiv \text{Op}(a \wedge b)$

Examples:

- a. John didn't see *anybody*. [John didn't see Mary or John didn't see Sue \equiv John didn't see (Mary and Sue)] [AM+]
- b. I don't think that *anybody* knows the answer. [I don't think that Lee knows the answer or I don't think that Mary knows the answer \equiv I don't think that (Lee and Mary) know the answer] [AM+]
- c. The bank avoided taking *any* decision. [The bank avoided taking decision A or the bank avoided taking decision B \neq The bank avoided taking (decision A and decision B)] [AM-] \mapsto [AA]

Within the anti-multiplicative area we check if the relevant operator is clausal negation.

(i) **Op** is clausal negation [D+]

Examples:

- a. John didn't see *anybody*. [D+] \mapsto [DN]
- b. I don't think that *anybody* knows the answer. [D-] \mapsto [AM]

Within the anti-additive non negative area we apply test (j) to distinguish free choice contexts.

(j) Test for free choice [FC+/-]: **Op**($a \vee \neg a$) is informative [FC+]

Examples:

- a. If you see *anybody*, you should tell me. [If you stay or go, you should tell me \mapsto antecedent is not informative] [FC-]
- b. John is taller than *anybody*. [John is taller than somebody or nobody \mapsto not informative] [FC-]
- c. You may kiss *anybody*. [You may stay or go \mapsto informative] [FC+] \mapsto [FC]

Within the non free choice contexts we distinguish the comparative constructions from the others.

(k) Comparative construction [CO+]

Examples:

- a. If you see *anybody*, tell me immediately. [CO-] \mapsto [CA]
- b. John is taller than *anybody*. [CO+] \mapsto [CO]

Further applications of the tests Consider now the following ambiguous example from Horn (2005:183):

(11) If she can solve *any* problem, she'll get a prize.

- a. ('existential') If there is any problem she can solve, ...
- b. ('universal') If she can solve every problem, ...

When applying our decision procedure to this example, at node (c) (the test for universal reading) we have to decide on what operator counts as the relevant **Op**. We have two candidates here: the conditional construction or the possibility modal *can*. In the first case (corresponding to the existential reading in (11a)) our terminal node will be **CA**, as illustrated in (12). In the second case, (corresponding to the universal reading in (11b)) our terminal node will be **FC**, as illustrated in (13):

- (12)
- a. If she can solve *any*_i problem, she'll get a prize. # It_i is a very difficult question. [S-]
 - b. If she can solve *any* problem, she'll get a prize. \Rightarrow For every problem x : (if she can solve x , then she'll get a prize) [V+]
 - c. If she solves problem A or problem B, she'll get a prize. \Rightarrow If she solves problem A, she'll get a prize and if she solves problem B, she'll get a prize. [AA+]
 - d. If she solves or doesn't solve a problem, she'll get a prize \mapsto antecedent is not inconsistent [Neg-]

- e. If she solves or doesn't solve a problem, she'll get a prize \mapsto antecedent is not informative [FC-]
 - f. If she can solve *any* problem, she'll get a prize. [CO-] \mapsto [CA]
- (13)
- a. If she can solve *any_i* problem, she'll get a prize. # It_i is a very difficult question. [S-]
 - b. If she can solve *any* problem, she'll get a prize \Rightarrow If (for every problem x : she can solve x), then she'll get a prize [V+]
 - c. She can solve problem A or problem B \Rightarrow She can solve problem A and she can solve problem B [AA+]
 - d. She can solve a problem or not \mapsto not inconsistent [Neg-]
 - e. She can solve a problem or not \mapsto informative [FC+] \mapsto [FC]

In ambiguous cases like this one, if the context did not disambiguate the intended reading, the sentences were annotated with both possible functions. To keep the randomly chosen occurrences stable the readings were counted as 0.5.

While these tests proved useful for many cases, there were examples for which our decision tree was inconclusive, and we conclude the section by discussing one of these cases. Consider the following example from Horn (2005), (see also Vlachou 2007):

- (14) I do not want to go to bed with just *anyone* anymore. I have to be attracted to them sexually.

Applying our tests for specific and for universal reading leads us to place this sentence in the non-specific existential area in our map. This area contains only two functions: Q and IR. Neither of these functions, however, are appropriate for this occurrence since, to quote Horn '*any* appears here in its free choice incarnation' (Horn 2005:185).

- (15)
- a. I do not want to go to bed with just *anyone_i* anymore. # He_i is very handsome. [S-]
 - b. I do not want to go to bed with just *anyone* anymore. [\nexists for every x : I don't want to go to bed with x] [V-]
 - c. I do not want to go to bed with just *anyone* anymore. [Q-], but not [IR] either.

To cover these cases we decided to introduce a new function, the indiscriminacy function IND. During annotation we have also introduced other off-map functions to label uses which were not strictly indefinite. One example is the *no-matter* function of which we give here an illustration in Czech:

- (16) A u jsme v kterkoli zemi, vude nachzme slun lidi.
 let already be:1PL in any country everywhere find:1PL polite people
 'No matter in which country you are, you can find polite people everywhere.'

In other cases where our decision tree was inconclusive, we left the issue open, and labeled the occurrence as unclear.

3 Some findings

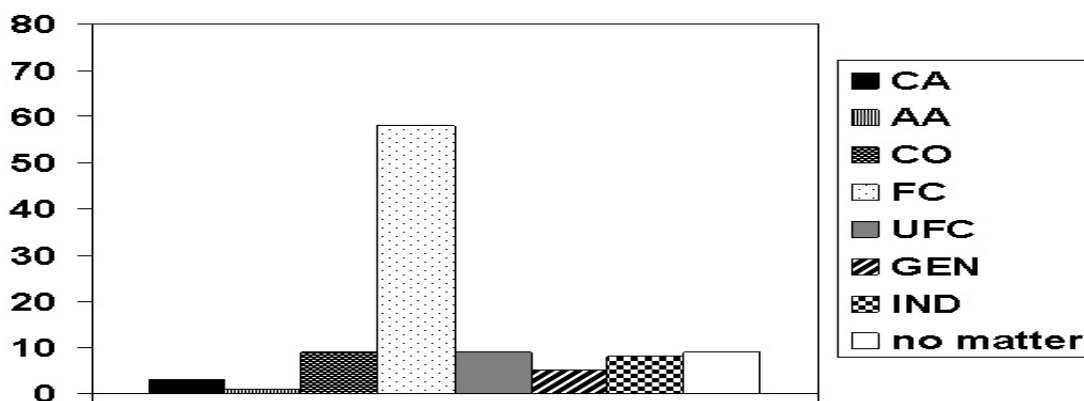
As an illustration of the results of the corpus studies we present the synchronic and the diachronic data of Spanish *cualquiera* and Dutch *wie dan ook*, two constructions that share the property of employing wh-morphology to express free choice meanings.

3.1 Spanish *cualquiera*

For the study of this item, we used *El Corpus del Español* created by Mark Davies. We randomly selected 100 occurrences of *cualquiera* from four sections, namely 1200s (7.9 millions of words), 1500s (19.7 millions of words), 1700s (11.5 millions of words), and 1900s (22.8 millions of words), which represent the four periods in which the history of Spanish has traditionally been divided (cf. Lapesa 1964). We used as a query the sequence *ualq*, which yielded all sorts of spelling variants of the item plus only ten instances of completely unrelated words, which were excluded.

Cualquiera (pronoun), or *cualquier* (determiner), composed of *cual* (‘which/who’) plus *quier(a)* (‘want:3.PRES.SUBJ’) has the following distribution in current Spanish:

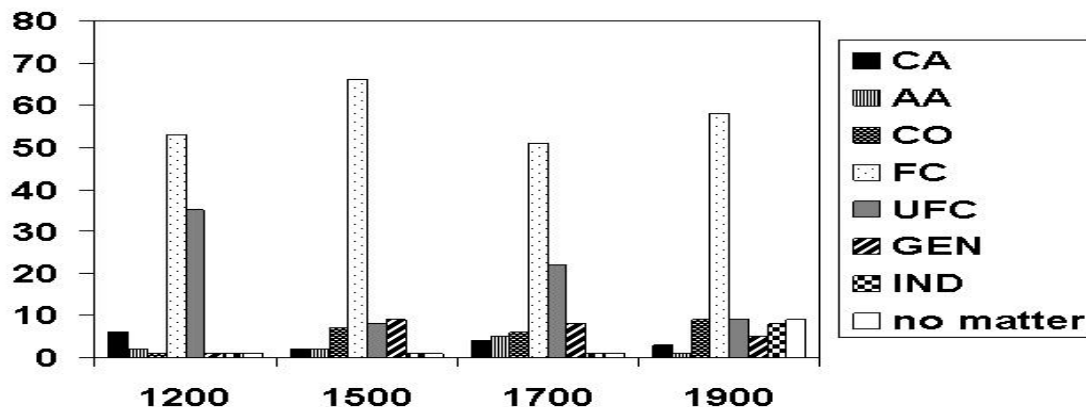
(17) Functions covered by *cualquiera* in current Spanish



This distribution, just like those of the other indefinites discussed in Aguilar-Guevara et al. (2010), confirms Haspelmath’s prediction that an indefinite always covers functions that are contiguous in the map.

Let us now discuss the historical development of *cualquiera*. This construction has been claimed to have emerged in Spanish as result of a grammaticalization process through which free relative clauses were reanalyzed as indefinite noun phrases (cf. Company-Company and Pozas-Loyo 2009). Presumably, this process has occurred in early stages of the history of Spanish and in consequence *cualquiera*, as a word, is already recurrently found in the first documentations of Spanish, which date back to the thirteenth century. As discussed in Aguilar-Guevara et al. (2010), the number of instances of *cualquiera* that were documented for each period studied suggest that the use of the construction is already consolidated quite early. The distribution of the functions that *cualquiera* covers throughout these periods points out to a similar conclusion:

(18) Functions covered by *cualquiera* in 1200s, 1500s, 1700s and 1900s



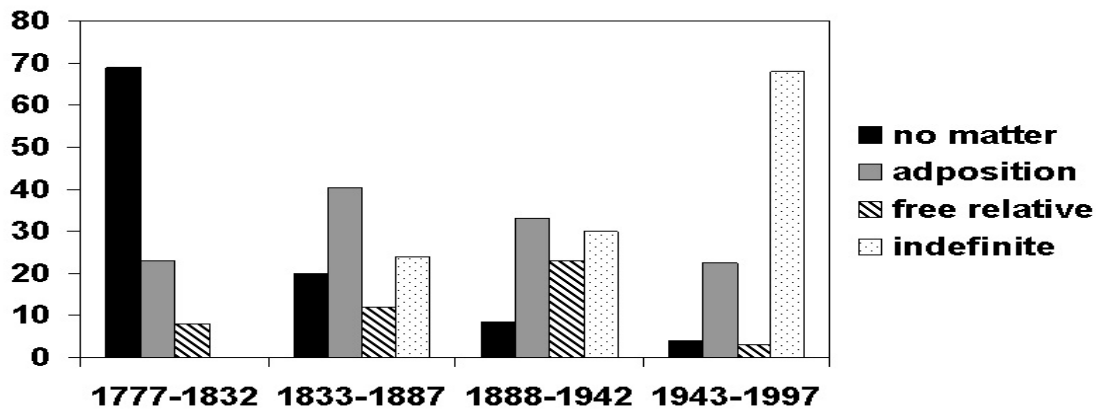
The most noteworthy observation about this distribution is that, generally speaking, it has remained pretty similar throughout the four periods. The FC function is clearly the most dominant since the first period, but some other functions contiguous in the map, namely, CA, CO and AA, as well as the functions UFC and GEN, have some presence as well. Interestingly, the UFC function displays a remarkable decrease as from the 1500s. In Aguilar-Guevara et al. (2010), we tentatively attribute this to the fact that *cualquiera*, as part of its grammaticalization, occurs less and less frequently accompanied by post-nominal modifiers such as restrictive relative clauses and prepositional adjuncts, which typically serve as licenser of free choice items in UFC uses (e.g. *John kissed any woman #(with red hair)*). The last important observation is that two more off-map functions, namely IND and *no-matter*, appear in the 1500s and gain presence by the 1900s. The late emergence of the *no-matter* function will turn particularly interesting in light of the development of the Dutch indefinite *wie dan ook*.

Given the early grammaticalization of *cualquiera* and stable distribution of its functions, we could not really attest much of the process this compound went through in order to behave as it does nowadays. This motivated us to study *wie dan ook*, an indefinite comparable to *cualquiera* in meaning and (partly) in form, but that emerged in Dutch more recently and that even in these days appears to be ‘less’ grammaticalized than *cualquiera*.

3.2 Dutch *wie dan ook*

The Dutch diachronic study, reported in de Vos (2010), consisted of the analysis of occurrences of *wie dan ook* (‘who also then’) in written Dutch historical corpora (CD-ROM Middelnederlands (270 texts before 1300), DBNL (4458 texts from 1170-2010)). The first occurrence found is from 1777; the period of this item’s existence has therefore been divided into four phases, each covering 55 years of the item’s evolution. The outcome shows that *wie dan ook* went through a four-staged process of grammaticalization:

(19) Four stages in grammaticalization of *wie dan ook*



Stage I The first phase in the grammaticalization of *wie dan ook* as an indefinite is formed by three forms of the *no matter*-function. Characteristic of types of *no matter* constructions is that the *wh dan ook* is not part of the main clause yet: they all consist of either a *wh*-clause and a main clause, or a *wh*-clause within a main clause, as illustrated as follows:

- (20) a. *Wie dan ook* naar het feest komt; ik zal blij zijn.
 ‘Whoever comes to the party; I will be happy.’
 b. [*Wie dan ook* naar het feest komt]_i; hij_i zal blij zijn.
 ‘[Whoever comes to the party]_i; he_i will be happy.’
 c. Jan, (of) *wie dan ook* hij mag zijn, zal blij zijn.
 ‘John, (or) whoever he may be, will be happy.’

These forms occur around the same time. Together, they seem particularly frequent in the first phase, forming a significant majority of the total amount of occurrences here, with this relative amount decreasing in the three phases that follow (cf. the black bars in graph (19)).

Stage II In the following stage in the development of *wie dan ook* as an indefinite, *no matter*-constructions are shortened to adpositions, thus getting one step closer to becoming a grammaticalized indefinite. Adpositions have the following form: [..., [*wie dan ook*], ...]. They are shortenings of the *no matter*-function, formed by the ellipsis of the predicate. Although they do not form a separate *wh*-clause next to or within a main clause anymore, they are still not part of the actual sentence and therefore no real indefinites: they merely modify the noun they are placed after.

- (21) Als er iemand_i, *wie dan ook*_i, naar het feest komt, zal ik blij zijn.
 ‘If someone, whoever/anyone, comes to the party, I will be happy.’

As the grey bars in (19) show, this adpositional modification with a *wie dan ook* (with ignorance or indifference meaning) is particularly frequent in the second phase in the development of this indefinite.

Stage III The third phase, the *free relative*-stage, shows a further integration of the *wie dan ook*-clause into the sentence, though still not a full integration either. The Free Relative (FR) function, the biggest part of the total amount of occurrences of *wie dan ook* now, forms another spinoff of the *no matter* construction. However, whereas *no matter*-sentences still form combinations of *wh*-clauses (*wie dan ook* + predicate) and a main clause, the FR-function is more integrated than that, with the “*wie dan ook* + predicate” not forming a separate clause, but an actual part of the main clause, typically the subject. Examples of the FR-function have the following form: [[*wie dan ook* + predicate](,) VP], as illustrated in (22):

- (22) Wie dan ook naar het feest komt, zal blij zijn.
'Whoever comes to the party(,) will be happy.'

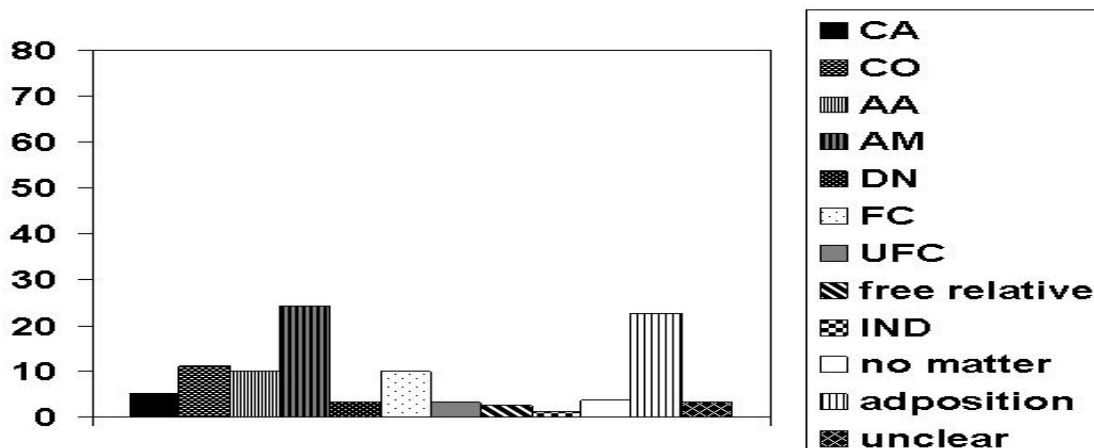
However, these subjects consisting of *wie dan ook* + predicate are often followed by a comma, thereby perhaps indicating that they are still seen as slightly standing outside of the actual sentence. Yet omitting the part starting with *wie dan ook* would give an incomplete thus ungrammatical sentence. This is a specific feature of the third phase; both the *no matter*-clauses and the adpositions can still be left out, of course sometimes causing a change in meaning of the sentence, but never with an incomplete sentence as a result. This shows how integrated a part of the sentence these occurrences of *wie dan ook* already form - although it apparently still feels a bit strange to the contemporary writer. Besides, these forms of *wie dan ook* are not as integrated yet as the plain indefinite will be.

Stage IV In this last stage of the grammaticalization of *wie dan ook*, the word group has finally become an indefinite. Examples of this kind form integrated parts of the sentence, with a plain *wie dan ook*, without any kind of predicate modifying it, being either subject or object: [. . . [*wie dan ook*] . . .].

- (23) Je mag wie dan ook uitnodigen voor het feest.
'You may invite anyone to the party.'

Indefinite uses of *wie dan ook* are attested from 1833 onwards, and their number increases in every phase, finally forming a vast majority of the occurrences in the fourth phase, as graph (19) illustrates. Here is the distribution of *wie dan ook* in stage IV:

(24) Functions covered by *wie dan ook* in stage IV (current Dutch)



Summarizing: Overall, what can be concluded is that the process of grammaticalization of *wie dan ook* as an indefinite roughly followed four stages, starting off as a *no-matter* construction in a separate *wh*-clause, slowly evolving into an adpositional modifier on its own, while also turning into a part of the main clause with predicate, eventually yielding to the true and plain indefinite *wie dan ook* as part of a sentence. Recall that the Spanish study showed a very late emergence of the *no-matter* function for *cualquiera*. This fact, combined with the phases of development of *wie dan ook*, constitutes evidence against unidirectionality in the acquisition of new functions: while the Dutch item was born with the *no-matter* function, the Spanish item starts its development from a free relative into a plain indefinite and only later allows the *no-matter* function to emerge.

Our initial hypothesis was that FC indefinites emerged as the result of a process of conventionalization of an originally pragmatic inference. The envisaged ‘conventionalization’ is in fact quite difficult to test because conversational implicatures are by definition not overtly expressed. The testing would have to consist in checking for a raising frequency of a conversational implicature of sentences with plain indefinites, then a development of a new morpheme which captures the implicature and then its grammaticalization. Alternatively, the morpheme that had already been used in the plain indefinite would change its function - the implicature would be built in. The latter is not what we observe. Yet, the described development of *wie dan ook* is consistent with the former scenario, with appositive *wie dan ook* as a new form which expresses the original implicature and later gets grammaticalized. More precisely, the grammaticalization path that we are describing for *wie dan ook* could be interpreted as a path from a conversational implicature, via a *conventional* implicature in the sense of Potts (2005)⁵ to a conventional meaning (i.e. core / at-issue semantics).

- (25) a. Jij mag iemand uitnodigen. (plain indefinite + conversational implicature)
 b. Jij mag iemand, *wie dan ook* (hij mag zijn), uitnodigen. (plain indefinite + conventional implicature)
 c. Jij mag *wie dan ook* uitnodigen (new FC indefinite)

⁵According to Potts (2005), appositives express conventional implicatures, i.e. not at-issue meanings.

To conclude, the emergence of *wie dan ook* as a plain indefinite counts as a classical example of grammaticalization, where the initial periphrastic usage of a *wh*-clause increased in frequency to such an extent that this usage got reanalyzed as being part of its lexical semantics. Such a process, as is often attested, takes place in a step-wise fashion. The adpositional usage results from the *no matter* usages of *wh*-clauses and can be taken to be the first lexicalization of a FC implicature. However, this adposition brings in new usage effects as well, such as its strong collocational distribution w.r.t. subjects and objects. This, in turn, then causes the next steps of the grammaticalization process: the replacement of DPs by the *wh*-element. Grammaticalization is thus not a big step from a lexical to a functional category (*in casu* from a *wh*-clause towards an indefinite), but a series of small steps, each possibly being the result of lexicalization of implicatures.

4 Conclusion

We have discussed the methodology adopted for a cross-linguistic synchronic and diachronic corpus study on free choice and epistemic indefinites. The study covered five indefinites in five languages. The main goal of the study was to verify the distribution of these indefinites on an extended version of Haspelmath's (1997) functional map, and to attest their historical development. One of the main conclusions of the synchronic studies was that there is no indefinite that violates the function contiguity. An interesting conclusion of the diachronic research was that the acquisition of new functions is not unidirectional. These studies could not confirm, but neither reject, our initial hypothesis on implicature fossilization.

References

- Ana Aguilar-Guevara, Maria Aloni, Angelika Port, Katrin Schulz, and Radek Šimík. Free choice items as fossils. Workshop on Indefiniteness Crosslinguistically (DGfS) Berlin, February 25/26, 2010.
- Luis Alonso-Ovalle and Paula Menéndez-Benito. Modal indefinites. *Natural Language Semantics*, 18:1–31, 2010.
- Concepción Company-Company and Julia Pozas-Loyo. Los indefinidos compuestos y los pronombres genérico-impersonales *omne* y *uno*. In Concepción Company-Company, editor, *Sintaxis histórica de la lengua española (Segunda parte: La frase nominal)*, pages 1073–1219. Fondo de Cultura Económica-Universidad Nacional Autónoma de México, México City, 2009.
- Veneeta Dayal. *Any* as inherently modal. *Linguistics and Philosophy*, 21:433–476, 1998.
- Machteld de Vos. *Wh dan ook*: The synchronic and diachronic study of the grammaticalization of a Dutch indefinite. BA thesis, University of Amsterdam, 2010.
- Anastasia Giannakidou. The meaning of free choice. *Linguistics and Philosophy*, 24:659–735, 2001.
- Martin Haspelmath. *Indefinite pronouns*. Oxford University Press, Oxford, 1997.
- Laurence Horn. Airport '86 revisited: Toward a unified indefinite *any*. In Gregory Carlson and Francis J. Pelletier, editors, *The Partee Effect*, pages 179–205. CSLI, Stanford, 2005.
- Jacques Jayez and Lucia Tovena. Epistemic determiners. *Journal of Semantics*, 23:217–250, 2006.
- Angelika Kratzer and Junko Shimoyama. Indeterminate pronouns: The view from Japanese. In

- Yukio Otsu, editor, *The proceedings of the Third Tokyo Conference on Psycholinguistics*, pages 1–25, Tokyo, 2002.
- Rafael Lapesa. *Historia de la lengua española*. Gredos, Madrid, 1964.
- Paula Menéndez-Benito. On universal Free Choice items. *Natural Language Semantics*, 18:33–64, 2010.
- Christopher Potts. *The Logic of Conventional Implicatures*. Oxford University Press, Oxford, 2005.
- Andreas van Cranenburgh, Raquel Fernandez, Katya Garmash, Marta Sznajder, and Maria Velema. Assessing the reliability of an annotation scheme for indefinites. Technical Report, ILLC, University of Amsterdam, 2011.
- Evangelia Vlachou. *Free Choice in and out of Context: The semantics and distribution of French, Greek and English Free Choice Items*. PhD thesis, LOT dissertation series 156, Utrecht, 2007.