Semantics and pragmatics of indefinites: methodology for a synchronic and diachronic corpus study

Ana Aguilar Guevara[1], Maria Aloni[2], Angelika Port[2], Radek Šimík[3], Machteld de Vos[2], and Hedde Zeijlstra[2]

[1]Utrecht University
[2]University of Amsterdam
[3]University of Potsdam

September 7, 2010

Abstract

The article discusses the methodology adopted for a cross-linguistic synchronic and diachronic corpus study on indefinites. The study covered five indefinite expressions in five languages: German *irgendein*, Italian *(uno) qualunque*, Spanish *cualquiera*, Dutch *wie dan ook* and Czech *kterýkoli*. The main goal of the study was to verify the distribution of these indefinites on the functional map of Haspelmath (1997), and to attest their historical development.

1 Theoretical Background

It is well known that the use of expressions with existential meaning (e.g. plain indefinites like English *somebody*, or German *jemand* 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, for instance as **free choice indefinites**[1]: Spanish *cualquiera*, Italian *qualunque*, Czech *kterýkoli*, Hungarian *akárki*, ..., and as **epistemic indefinites**[2]: Russian *to*-series, Czech *si*-series, German, *irgend*-series, Spanish *algun*, ...

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 grammaticalization (or fossilization) of an originally pragmatic inference:

> It may not be impossible for what starts life, so to speak, as a conversational implicature to become conventionalized. (Grice 1975:58)

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

(3) *plain indefinite (German)*
   a. **Jemand** hat angerufen.
      somebody has called
   b. Conventional meaning: Someone called
   c. Ignorance implicature: Speaker does not know who

(4) *epistemic indefinite (German irgend-)*

[1]E.g. Dayal (1998), Giannakidou (2001), Menéndez-Benito (2010).
[2]E.g. Alonso-Ovalle and Menéndez-Benito (2010), Jayez and Tovena (2006), Kratzer and Shimoyama (2002).

a. **Irgendjemand** hat angerufen.
somebody:UNKNOWN has called
b. Conventional meaning: Someone called and I do not know who

In languages with distinctive free choice 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, a number of cross-linguistic synchronic and diachronic studies have been combined in order to substatiate 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 *irgendein*, Czech *kterýkoli*, Spanish *cualquiera*, Dutch *wie dan ook*, and Italian *(uno) qualunque*. 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 how it might had happened diachronically (cf. intentionally left blank). In the diachronic corpus research we studied the historical development of the first three indefinite forms: Spanish *cualquiera*, German *irgendein* and Dutch *wie dan ook* (cf. intentionally left blank).

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

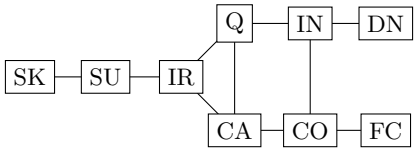
2 Corpus study: diagnostics and methodology

The methodology for these empirical studies is a form of functional labelling which combines both context (syntax) and meaning (semantics), using as a starting point the functions identified by Haspelmath (1997).

2.1 Haspelmath semantic map

Haspelmath's (1997) typological survey identified 9 main functions (context/meaning) for indefinite forms organized in an implicational map. Haspelmath proposes that an indefinite will always express a set of functions that are contiguous on the map. One prediction is that items which acquire new functions will develop first those functions that are adjacent to the original function.

(7) *Haspelmath's map*

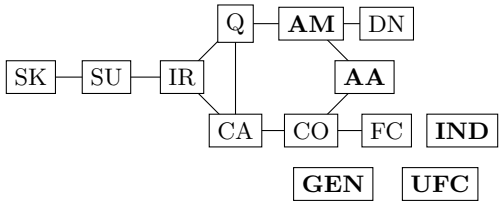


(8) *Functions on the map*

	Abbr	Label	Example
a.	SK	specific known	*Somebody* called. Guess who?
b.	SU	specific unknown	I heard *something*, but I couldn't tell what it was.
c.	IR	irrealis	You must try *somewhere* else.
d.	Q	question	Did *anybody* tell you anything about it?
e.	CA	conditional antecedent	If you see *anybody*, tell me immediately.
f.	CO	comparative	John is taller than *anybody*.
g.	IN	indirect negation	I don't think that *anybody* knows the answer.
h.	DN	direct negation	John didn't see *anybody*.
i.	FC	free choice	You may kiss *anybody*.

We have extended Haspelmath's original map as follows: the indirect negation function has been split into an antimorphic (AM) and an anti-additive (AA) function (cf. Zwarts 1998); and three new functions have been introduced contiguous to the free choice area, namely the indiscriminacy function (IND), the generic function (GEN) and the universal free choice (UFC) function. The precise collocation on the map of the latter three functions is still a matter of investigation.

(9) *Our extended map*



Motivation for this extension came from comparing more in detail the different items cross-linguistically. For example, German *irgend*-indefinites do not exhibit a generic function or a universal free choice function, but exibit a FC function, Dutch *wie dan ook* doesn't exhibit the generic function, but exhibits UFC and FC, whereas Italian *qualunque*, Czech *kterýkoli* and Spanish *cualquiera* exhibit all three functions. The new functions we added are marked with a → in the following illustration:

(10) *Functions on the map*

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

During annotation we have also introduced a number of 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:

(11) Ať už jsme v kterékoli zemi, všude nacházíme slušné 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.'

Off-map functions played a central role in the diachronic research discussed in section 4.

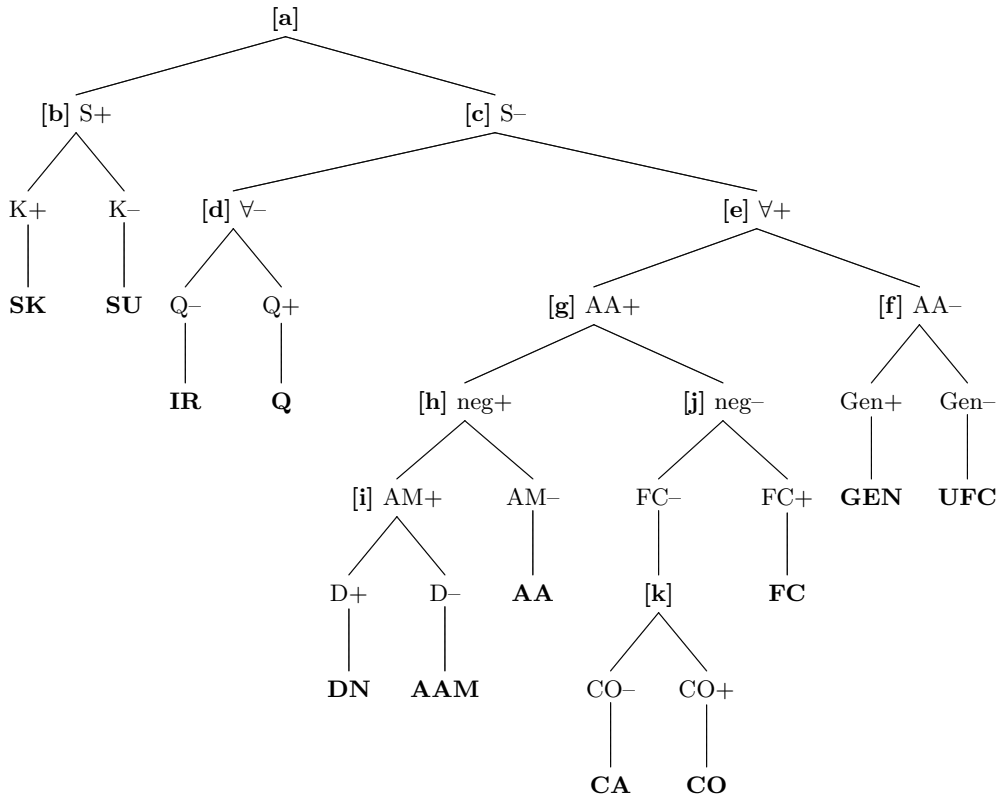
2.2 Methodology for semantic annotation

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. (12a); and *some* does not exhibit the comparative function CO because it does not have a universal meaning specified by the comparative function, cf. (12b).

(12) 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.'

In this section we discuss the tests we used to decide which function an indefinite exhibits in a certain context. These tests and the order in which they were applied are schematized in the following decision tree. The IND function and the off-map functions are not represented in this tree, the former will be shortly discussed at the end of the section.

(13) *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 (10) to exemplify our functional labels.

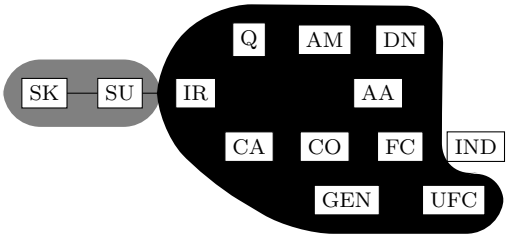
(a) Test for specificity [S+/–]:

Sentence (S): . . . indefinite$_i$ . . . **Possible Continuation (PC)**: She$_i$/he$_i$/it$_i$ . . . [S+]

Examples:

a. *Somebody*$_i$ called. She$_i$ wanted a new appointment. [S+]
b. I heard *something*$_i$. It$_i$ was very loud. [S+]
c. You must try *somewhere*$_i$ else. # It$_i$ is a very nice place. [S–]
d. Did *anybody*$_i$ tell you anything about it? # He$_i$ is a real chatterbox. [S–]
e. If you see *anybody*$_i$, tell me immediately. # He$_i$ is a nice guy. [S–]
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+] ↦ [**SK**]
b. I heard *something*, but I couldn't tell what it was. # Guess what? [K–] ↦ [**SU**]

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

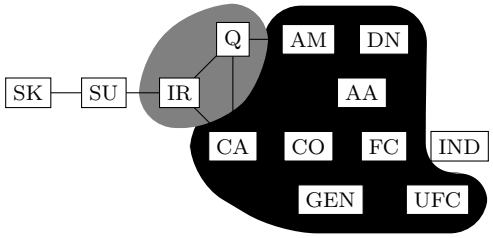
(c) Test for universal meaning [∀+/–]:

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

Examples:

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

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.

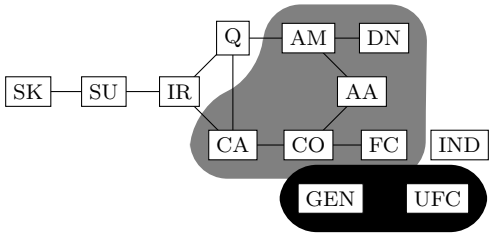
(e) Test for anti-additivity [AA+/–]: **Op**$(a \vee b) \Rightarrow$ **Op**$(a) \wedge$ **Op**(b) [AA+]

Examples:

a. If you see *anybody*, tell me immediately. [If you see John or Mary, tell me immediately ⇒ If you see John, tell me immediately and if you see Maria, tell me immediately] [AA+]
b. John is taller than *anybody*. [John is taller than Lee or Mary ⇒ 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 ⇒ 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 ⇒ 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 ⇒ 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 ⇒ 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 ≢ John kissed a woman with red hair [Gen–] ↦ [**UFC**]
b. *Any* dog has four legs ≡ A dog has four legs [Gen+] ↦ [**GEN**]

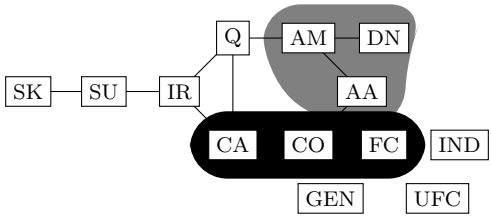
Within the anti-additive area we apply test **(g)** to distinguish negative contexts from non negative ones.

(g) Test for negative meaning [Neg+/–]: **Op**$(a \vee \neg a)$ is inconsistent [Neg+]

Examples:

a. John didn't see *anybody*. [John did not see somebody or nobody ↦ inconsistent] [Neg+]
b. I don't think that *anybody* knows the answer. [I don't think that the door is open or closed ↦ inconsistent] [Neg+]
c. The bank avoided taking *any* decision. [The bank avoided taking some decision or no decision] ↦ inconsistent] [Neg+]
d. You may kiss *anybody*. [You may kiss somebody or nobody ↦ not inconsistent] [Neg–]
e. If you see *anybody*, tell me. [If somebody or nobody comes, tell me ↦ not inconsistent] [Neg–]
f. John is taller than *anybody*. [John is taller than somebody or nobody ↦ 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: **Op**$(a)\vee$**Op**(b) ≡ **Op**$(a \wedge b)$

Examples:

a. John didn't see *anybody*. [John didn't see Mary or John didn't see Sue ≡ 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 ≡ 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 ≢ The bank avoided taking (decision A and decision B)] [AM–] ↦ **[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+] ↦ **[DN]**

b. I don't think that *anybody* knows the answer. [D–] ↦ **[AAM]**

Within the anti-additivive 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*, tell me immediately [If John comes or doesn't come, I will go to the party ↦ antecedent is not informative] [FC–]
b. John is taller than *anybody* [John is taller than somebody or nobody ↦ not informative] [FC–]
c. You may kiss *anybody* [You may go or stay ↦ informative] [FC+] ↦ **[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–] ↦ **[CA]**
b. John is taller than *anybody*. [CO+] ↦ **[CO]**

Further applications of the tests

Consider first example (14):

(14) Every door that *anyone* painted was locked.

As illustrated in (15), our battery of tests places these downward entailing contexts in the CA area rather than in the negative area (AA, AAM, DN), showing that the restrictor of an universal and the antecedent of a conditional share the same semantic properties, as implicitly assumed by most of the modal analyses of conditional constructions.

(15)
a. Every door that *anyone*$_i$ painted was locked. # He was very quick. [S–]
b. Every door that *anyone* painted was locked ⇒ for all x: every door that x painted was locked [∀–]
c. Every door that Maria or Lee painted was locked ⇒ Every door that Maria painted was locked and every door that Lee painted was locked [AA+]
d. Every door that Mary painted or didn't paint was locked ↦ restriction not inconsistent [Neg–]
e. Every door that Mary painted or didn't paint was locked ↦ restriction not informative [FC–] ⇒ [**CA**]

Next consider the following ambiguous example from (Horn 2005:183):

(16) 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 (16)-a) our terminal node will be **CA**, as illustrated in (17). In the second case, (corresponding to the universal reading in (16)-b) our terminal node will be **FC**, as illustrated in (18):

(17)
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. ⇒ For every problem x: (if she can solve x, then she'll get a prize) [∀+]
c. If she solves problem A or problem B, she'll get a prize. ⇒ 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 ↦ antecedent is not inconsistent [Neg–]
e. If she solves or doesn't solve a problem, she'll get a prize ↦ antecedent is not informative [FC–]
f. If she can solve *any* problem, she'll get a prize. [CO–] ↦ [**CA**]

(18)
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 ⇒ If (for every problem x: she can solve x), then she'll get a prize [∀+]

c. She can solve problem A or problem B ⇒ She can solve problem A and she can solve problem B [AA+]
d. She can solve a problem or not ↦ not inconsistent [Neg–]
e. She can solve a problem or not ↦ informative] [FC+] ↦ [**FC**]

In ambiguous cases like this one, if the context did not disambiguate the intended reading, the sentences were annotated with both possible functions.[3]

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):

(19) 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).

(20) a. I do not want to go to bed with just *anyone*$_i$ anymore. # He$_i$ is very handsome. [S–]
b. I don't want to go to bed with just *anyone* anymore. [$\not\Rightarrow$ for every$x$: I don't want to go to bed with x] [∀–]
c. I don't 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, discussed above. In other cases where our decision tree was inconclusive, we left the issue open, and labeled the occurrence as unclear.

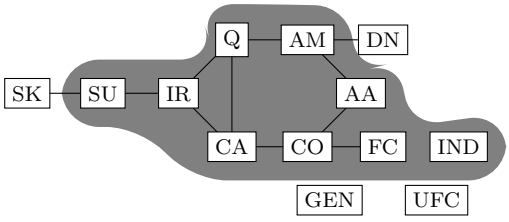
3 Results synchronic corpus study

As an illustration of our results this section compares the attested (percentage) distribution of the functions for German *irgendein* and Czech *kterýkoli*.

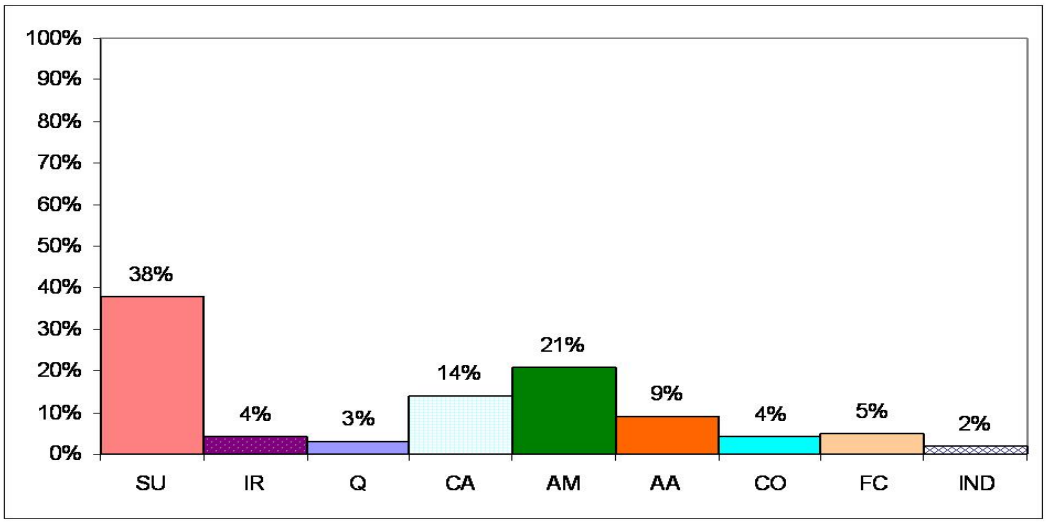
German

- Item: *irgendein* [*irgend* + *ein* 'a']
- Corpus: DWDS (Berlin-Brandenburgische Akademie der Wissenschaften; 100 million tokens, written, various registers)
- Query: irgendein* [six possible forms: *irgendein, irgendeine, irgendeiner, irgendeines, irgendeinen, irgendeinem*]
- Number of occurrences: 5975 out of which 4835 available (due to copyright)
- Labeled: 300 random occurrences

(21) *Distribution*



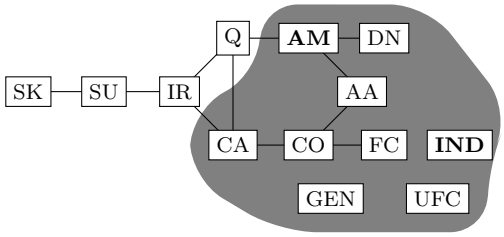
[3]To keep the randomly chosen occurrences stable the readings were counted as 0.5.

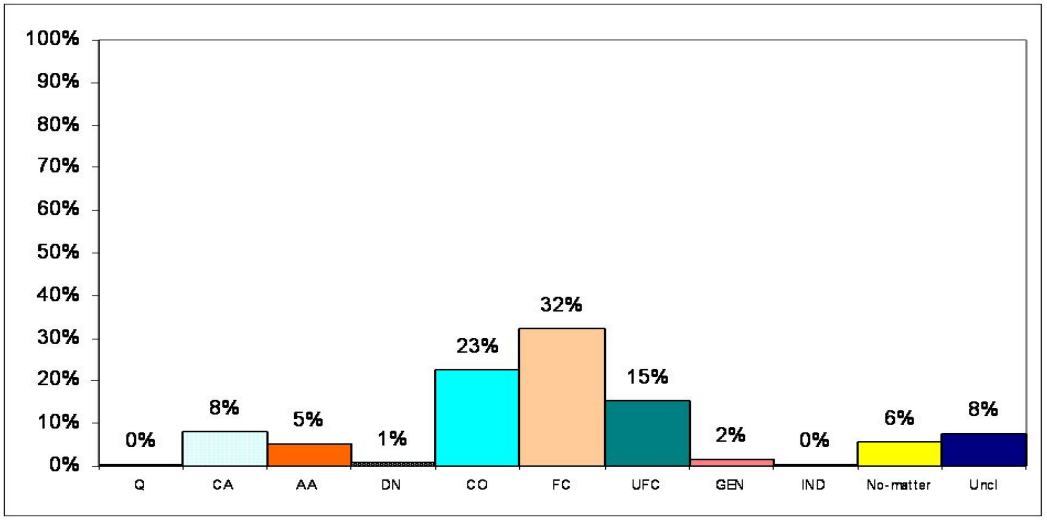


Czech

- Item: *kterýkoli* [*který* 'which' + *koli*; *li* is a particle used in yes/no questions and conditionals in present Czech]
- Corpus: Český národní korpus ČNK (Czech national corpus); subcorpus: SYN (synchronic corpus); URL http://korpus.cz/corpora/
- Query: kterýkoli [22 forms: 6 grammatical cases / 6 noun classes / capital/small initial letters]
- Number of occurrences: 7843
- Labeled: 300 random occurrences

(22) *Distribution*





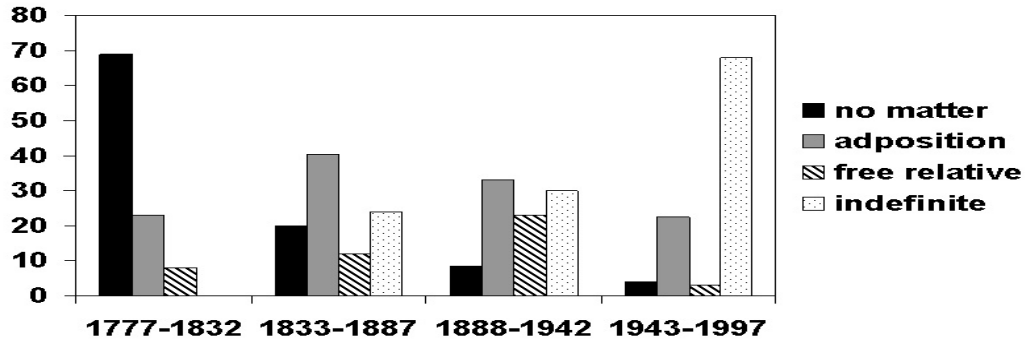
Discussion The main prediction of Haspelmath were confirmed by our synchronic corpus research: there is no indefinite that violates the function contiguity. Further our study attested significant differences in meaning and distribution between the studied items which are often considered to belong to the same class of free choice indefinites in the linguistic literature (cf. Giannakidou 2009).

4 Diachronic research

In the diachronic research we studied the historical developments of German *irgendein*, Dutch *wie dan ook* and Spanish *cualquiera* (cf. intentionally left blank). As an illustration of our results we will discuss in this section an interesting asymmetry we have attested in the development of the Dutch and Spanish indefinites.

The first found occurrence of Dutch *wie dan ook* 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 of what originally were implicatures, thus confirming the hypothesis that these semantics effects are indeed the result of grammaticalization of Gricean implicatures, and showing as well, that such processes of grammaticalization underwent a particular pathway before the construction allowed indefinite usages indicated in the graph (23) and examples in (24) below:

(23) *Four phases in the development of Dutch 'wie dan ook'*



(24) Attested grammaticalization process for *wie dan ook*:

a. *No-matter*: [WH dan ook + predicate], [main clause]
Wie dan ook naar het feest komt, ik zal blij zijn.
'Whoever may come to the party, I will be happy'

b. *Adposition*: [. . . , [WH dan ook], . . .]
Als er iemand$_i$, **wie dan ook**$_i$, naar het feest komt, zal ik blij zijn.
'If someone$_i$, whoever/anyone$_i$, comes to the party, I will be happy'

c. *Free Relative*: [[WH dan ook + predicate](,) VP]
Wie dan ook naar het feest komt, zal blij zijn.
'Whoever comes to the party will be happy'

d. *Indefinite*: [. . . [WH dan ook] . . .]
Je mag **wie dan ook** uitnodigen voor het feest.
'You may invite anyone to the party'

The actual grammaticalization of *cualquiera* instead was not attested in our corpus study. Our hypothesis is that it occurred before the time covered by available corpora. Indeed there is evidence that *cualquiera*, just like other indefinite compounds in Spanish (*quiquier, quequier, quien quiera*), was born in the language as a result of grammaticalization processes and not just as calques of Latin impersonal indefinites (cf. Company-Company and Pozas-Loyo 2009):

(25) Hypothesized grammaticalization process for *cualquiera*

a. *Free relative clause*
Haga en él **cual** castigo **quiera**.
do on him which punishment want:3.PRES.SUBJ

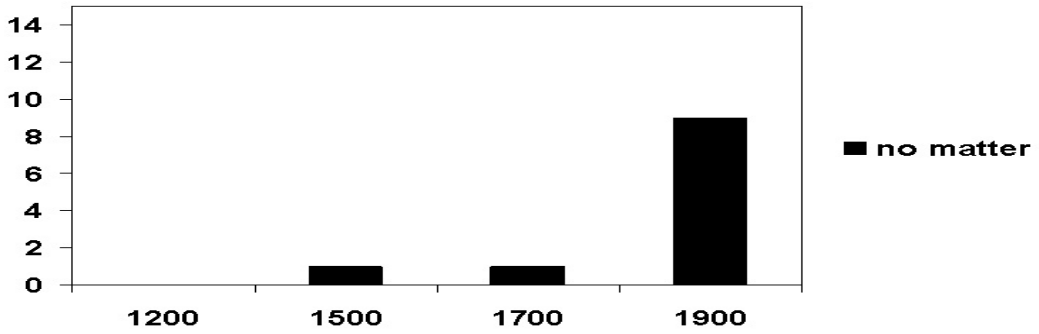
b. *Phrasal compound*
Haga en él **cual** **quiera** castigo.
do on him which want:3.PRES.SUBJ punishment

c. *Indefinite*

Haga en él **cualquier(a)** castigo
do on him whichever punishment

In the corpus study, 50-150 occurrences of *cualquier(a)* have been labeled for each diachronic stage: 1200s, 1500s, 1700s, 1900s. These centuries represent the periods in which the history of Spanish has been divided (cf. Lapesa 1964). One interesting result of this study was the late emergence of the *no-matter* function.

(26) *Late emergence of no-matter function for Spanish '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.

5 Conclusion

The article presents 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 the functional map of Haspelmath (1997), and to attest their historical development. Taking Hasplemath's functions as a starting point, a number of randomly selected occurrences of each of our items have been syntactically and semantically annotated (context/meaning). One of the main conclusions of the synchronic studies is that there is no indefinite that violates the function contiguity. The main conclusion of the diachronic studies is that the acquisition of new functions is not unidirectional. In this article we focused on the battery of tests that we have used during the annotation.

References

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 le 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.

Anastasia Giannakidou. The meaning of free choice. *Linguistics and Philosophy*, 24:659–735, 2001.

Anastasia Giannakidou. Negative and positive polarity items: licensing, compositionality and variation. Manuscript prepared for C. Maienborn and K. von Heusinger and P. Portner, editors, *Semantics: An International Handbook of Natural Language Meaning*, Mouton de Gruyter, Berlin, 2009.

H. Paul Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and semantics, Vol 3: Speech acts*, pages 41–58. Seminar Press, New York, 1975.

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

Laurence R. Horn. Airport '86 revisited: Toward a unified indefinite *any*. In G. Carlson and F.J. Pelletier, editors, *The Partee Effect*, pages 179–205. CSLI, Stanford, 2005.

Jacques Jayez and Lucia M. Tovena. Epistemic determiners. *Journal of Semantics*, 23 (3):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.

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.

Frans Zwarts. Three types of polarity. In F. Hamm and E. Hinrichs, editors, *Plurality and quantification*, pages 177–238. Kluwer, Dordrecht, 1998.