Indefinites as fossils: a synchronic and diachronic corpus study

Ana Aguilar-Guevara[1], Maria Aloni[2], Angelika Port[2], Radek Šimík[3], Machteld de Vos[4], and Hedde Zeijlstra[2]

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

December 13, 2010

Abstract

The article reports on 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. On Haspelmath's proposal an indefinite will always express a set of functions that are contiguous on the map. A further prediction is that, if an indefinite acquires new functions, it will develop first those that are adjacent to the original function. The main result of the synchronic studies is that none of the indefinites investigated violates the function contiguity. The main conclusion of the diachronic studies is that the acquisition of new functions is not unidirectional.

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, such as **free choice indefinites**[1]: Italian *-unque*-series, Czech *koli*-series, Dutch *dan ook*-series, Spanish *cualquier*-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.

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

---

[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).

(3)   *Plain indefinite (German)*

    a.   **Jemand** hat angerufen.
       somebody has called

    b.   Conventional meaning: Someone called

    c.   Ignorance implicature: The speaker does not know who

(4)   *Epistemic indefinite pronoun (German 'irgendjemand')*

    a.   **Irgendjemand**   hat angerufen.
       somebody:UNKNOWN has called

    b.   Conventional meaning: Someone 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)   *Free choice 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 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 had 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 discuss the methodology developed 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

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 for the identification of the relevant categories was Haspelmath's (1997) implicational map. In this section, we introduce Haspelmath's map, we extend it for the purpose of a more detailed NPI/FC classification, and we provide an explicit set of logico-semantic criteria, according to which indefinite pronouns are assigned functions on the map.

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 (where two functions are contiguous iff they are connected by a line). 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*

---
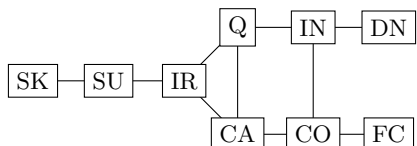
[3]An assessment of the methodology (by measuring inter-annotator agreement) has been planned for January 2011.

(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*.

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.   Somebody /# anybody called.
       b.   Berlin is bigger than any /# some Czech city.
             'For all Czech cities it holds that Berlin is bigger than they are.'

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 two new functions have been introduced contiguous to the free choice area, namely the generic function (GEN) and the universal free choice (UFC) function. The precise placement on the map (i.e. connecting lines determining function contiguity) of the latter two functions is still a matter of investigation.

(10)   *Our extended map*
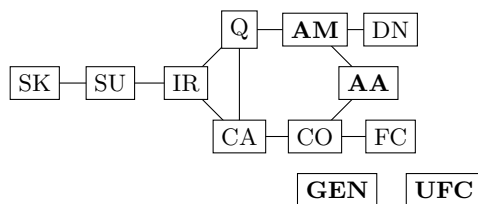
Motivation for this extension came from comparing more in detail the different items cross-linguistically. For example, German *irgend*-indefinites do not exhibit a generic function or a universal free choice function, but exibit a FC function, Italian *uno qualunque* doesn't exhibit the universal free choice function, but does exhibit FC and GEN, 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:

(11)   *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.

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:
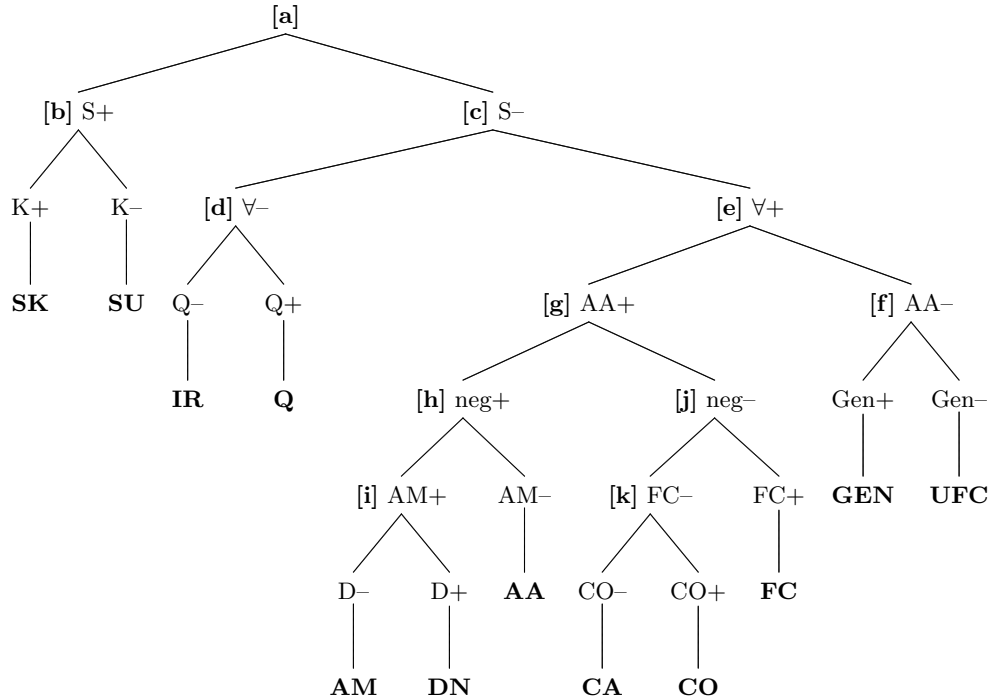
(12)   Ať už   jsme   v kterékoli zemi,   všude   nacházíme slušné lidi.
        let already be:1PL in any   country everywhere find:1PL   polite people
        'No matter in which country you are, you can find polite people everywhere.'

3

The *no-matter* function played a central role in the diachronic research discussed in section 4.

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. Off-map functions are not represented in this tree.

(13)   *Decision tree*

```
                                    [a]
                        ┌────────────┴────────────┐
                    [b] S+                      [c] S−
                  ┌────┴────┐          ┌──────────┴──────────┐
                 K+         K−      [d] ∀−                 [e] ∀+
                  │          │     ┌──┴──┐          ┌─────────┴─────────┐
                 SK         SU    Q−    Q+      [g] AA+              [f] AA−
                                   │     │     ┌────┴────┐         ┌────┴────┐
                                  IR     Q  [h] neg+   [j] neg−  Gen+     Gen−
                                            ┌───┴───┐   ┌───┴───┐   │        │
                                        [i] AM+   AM− [k] FC−  FC+  GEN     UFC
                                        ┌──┴──┐    │  ┌──┴──┐   │
                                       D−    D+   AA CO−  CO+   FC
                                        │     │      │     │
                                       AM    DN     CA    CO
```

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 (11) to exemplify our functional labels.

(a)   Test for specificity [S+/−]:

Sentence (S): . . . indefinite$_i$ . . . **Possible Continuation (PC)**: . . . pronoun$_i$ . . .   [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).
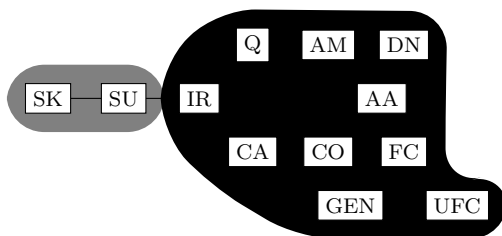
Within the specific area we apply test **(b)** to distinguish the specific known from the specific unknown function.

(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 [∀+/−]:

. . . **Op** (. . . indefinite . . .) . . . ⇒ . . . ∀x (**Op**. . . x . . .) . . .

Examples:

a. You must try *somewhere* else ⇏ for all places x: you must try x   [∀−]
b. Did *anybody* tell you anything about it? ⇏ 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).
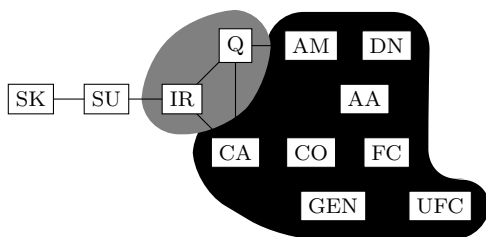
Within the existential area we distinguish polar questions from irrealis non-specific constructions via step **(d)**.

(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 ∨ b) ⇒ **Op**(a)∧**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 ⇒ If you see John, you should tell me immediately and if you see Maria, you should tell me immediately]   [AA+]

5

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 ⇏ John kissed Lee and John kissed Bea] [AA–]

h. *Any* dog has four legs. [Fido or Bobby has four legs ⇏ 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).
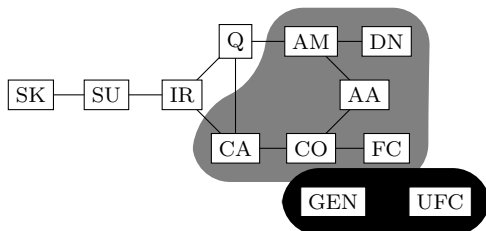
Within the non anti-additive area we apply test **(f)** to distinguish generic from universal free choice readings.

(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 didn't stay or go ↦ 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 being open or closed] ↦ inconsistent] [Neg+]

d. You may kiss *anybody*. [You may stay or go ↦ not inconsistent] [Neg–]

e. If you see *anybody*, you should tell me. [If you stay or go, you should 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).
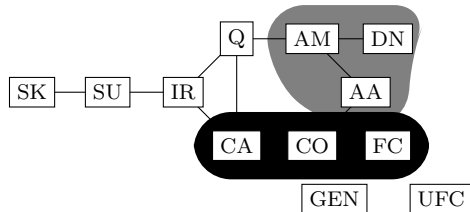
Within the negative area we apply test **(h)** to distinguish anti-multiplicative contexts from plain negative ones.

(h) Test for anti-multiplicativity: **Op**$(a) \vee$ **Op**$(b) \equiv$ **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–] ↦ **[AM]**

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

(j)   <u>Test for free choice [FC+/–]</u>: **Op**(a ∨ ¬a) is informative   [FC+]

     Examples:

   a.   If you see *anybody*, you should tell me. [If you stay or go, you should tell me ↦ 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 stay or go ↦ 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

Example one   Consider example (14) with *anyone* occurring in the restriction of a universal quantifier:

(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, AM, DN), showing that the restrictor of a 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]**

Example two   Consider now sentence (16), which has received some attention in the literature on free choice (Menéndez-Benito 2010: p.54–55, ex.83).

(16)   John must answer any question.

As illustrated in (17), the application of our tests leads us to classify the occurrence of *any* in (16) as a universal free choice use rather than a generic use:

(17)   a.   John must answer any question$_i$. # *It*$_i$ is very difficult.   [S–]

       b.   John must answer any question ⇒ for every question x: John must answer x [∀+]

       c.   John must answer question a or question b ⇏ John must answer question a and John must answer question b   [AA–]

    d.   John must answer any question. $\not\equiv$ John must answer a question.   [Gen–] $\mapsto$ [**UFC**]

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

(18)   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 (18a)) our terminal node will be **CA**, as illustrated in (19). In the second case, (corresponding to the universal reading in (18b)) our terminal node will be **FC**, as illustrated in (20):

(19)   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)   [$\forall$+]
       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**]

(20)   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   [$\forall$+]
       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.

Example four   Consider the following example from Horn (2005), (see also Vlachou 2007):

(21)   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).

(22)   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. [$\not\Rightarrow$ for every x: I don't want to go to bed with x]   [$\forall$–]
       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. 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*, Czech *kterýkoli*, Spanish *cualquiera*, Italian *(uno) qualunque* and Dutch *wie dan ook*.

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

(23)   *Distribution*
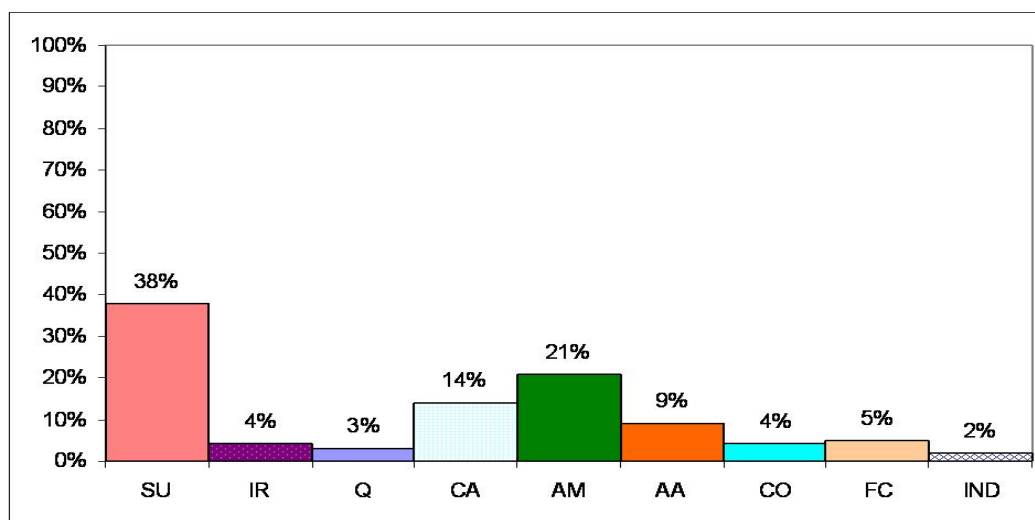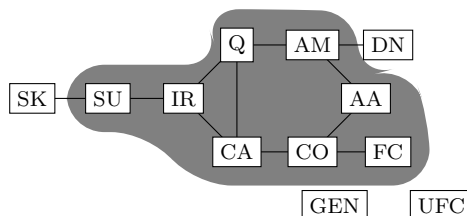
Czech

- Item: *kterýkoli* [*který* 'which' + *koli*; *li* is a particle that can mark (in the form of a verbal affix) embedded yes/no questions and conditional clauses, though only in a formal register]
- 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

(24)   *Distribution*
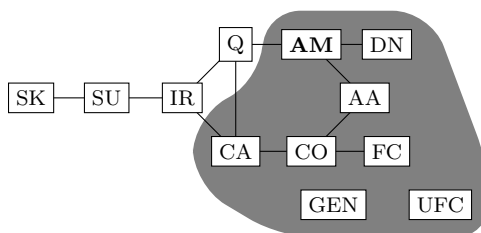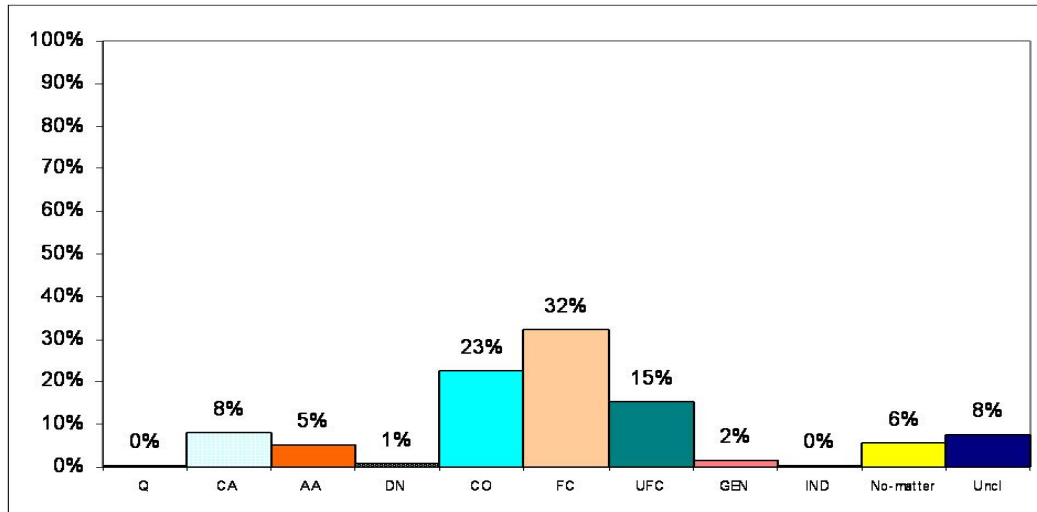
Spanish

- Item: *cualquiera* [*cual-* 'which' + *quiera* 'want:PRES.SUBJ.3']
- Corpus: CORPUS DEL ESPAÑOL (by Mark Davies; 100 million words)
- Query: *ualq* [all possible forms of *cualquier(a)* + 10 instances of completely unrelated words, which were excluded]
- Number of occurrences: 7744
- Labeled: 200 random occurrences

(25)   *Distribution*
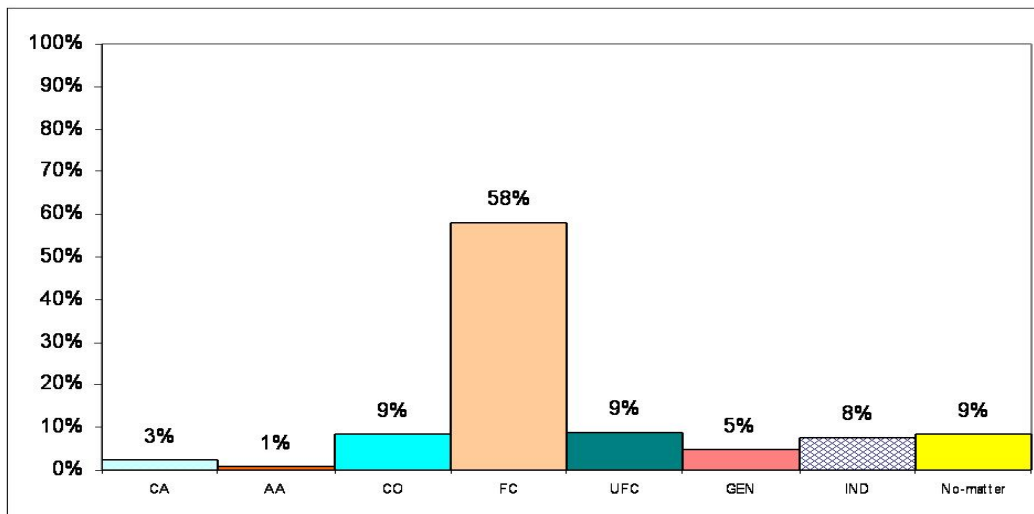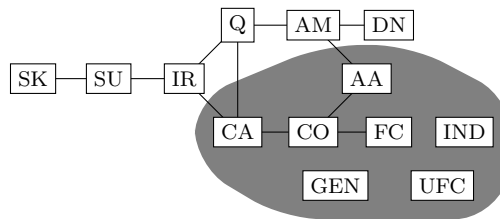
Italian

- Item: *qualunque* [from Latin *qualiscúmque* composed from *qualis* 'what' + *cúmque* 'ever']
- Corpus: CORIS (100 million words; various registers; a synchronic corpus of written language, whose component texts belong, roughly speaking, to the 1980s and 1990s, with a somewhat wider temporal collocation as far as narrative is concerned)
- Number of occurrences: 7591
- Labeled: 300 random occurrences

(26)   *Distribution*
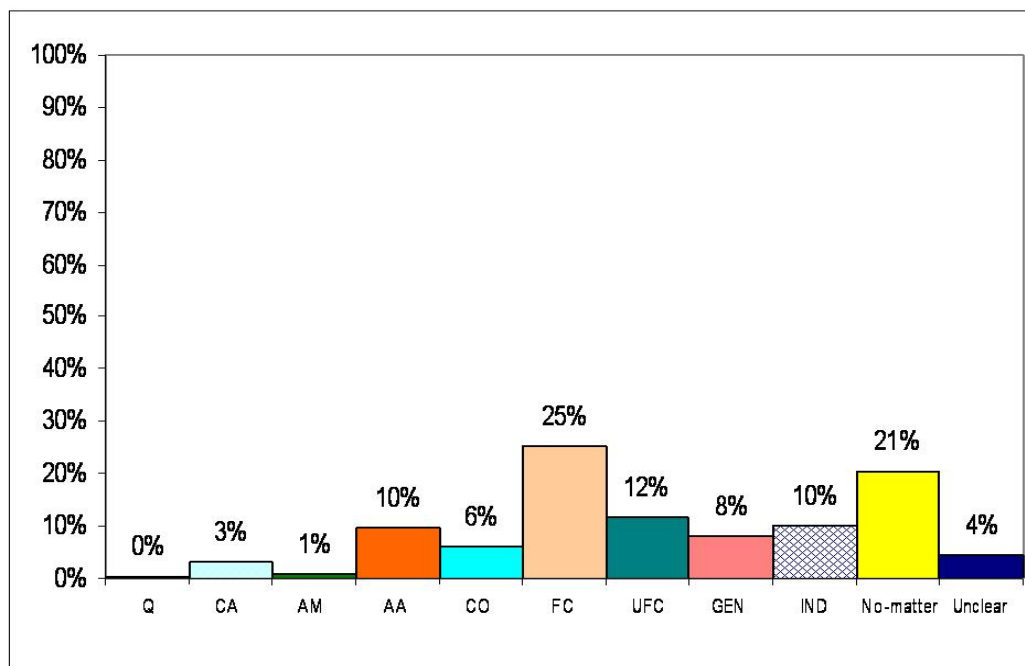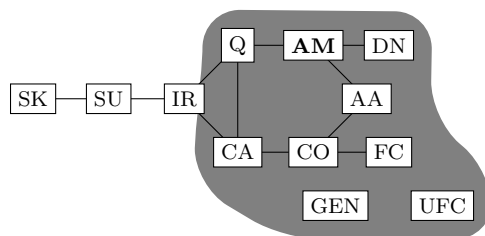
- *Qualunque* can occur in these forms:

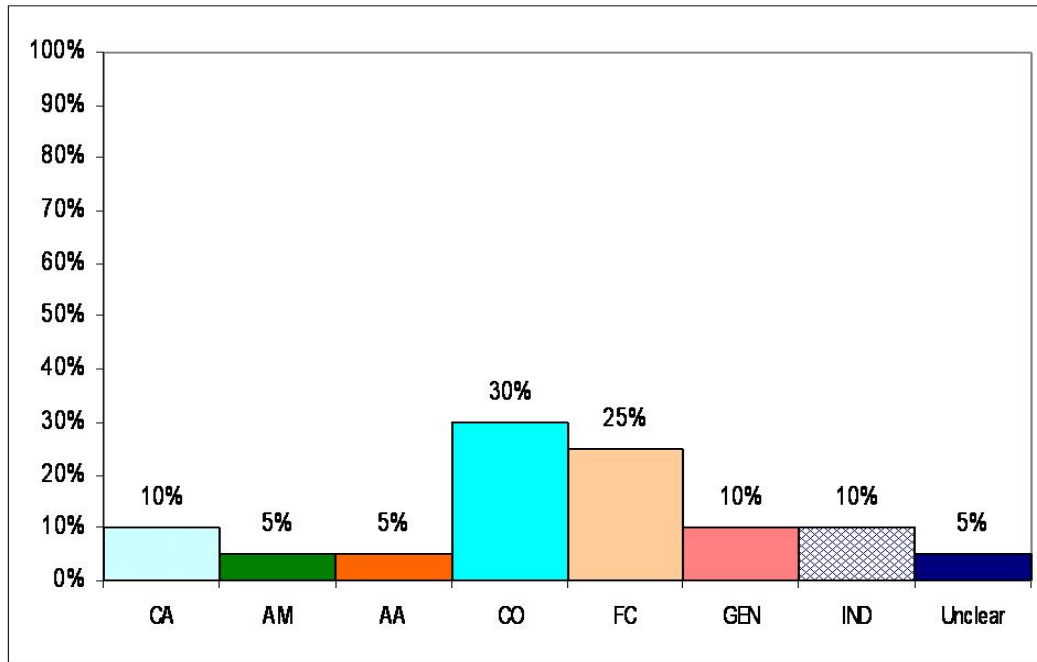| (27)	a.	Plain determiner: qualunque + N	184 (61,33%)
	b.	Existential determiner (ExD): un(a/o) + qualunque + N	20 (6,66 %)
	c.	PostN (probably Adj): Det + N + qualunque	32 (10,66 %)
	d.	Unclear PostN or ExD:	2 (0,68 %)
	e.	In no-matter constructions:	62 (20,66 %)

Qualunque + N

(28)   *Distribution*
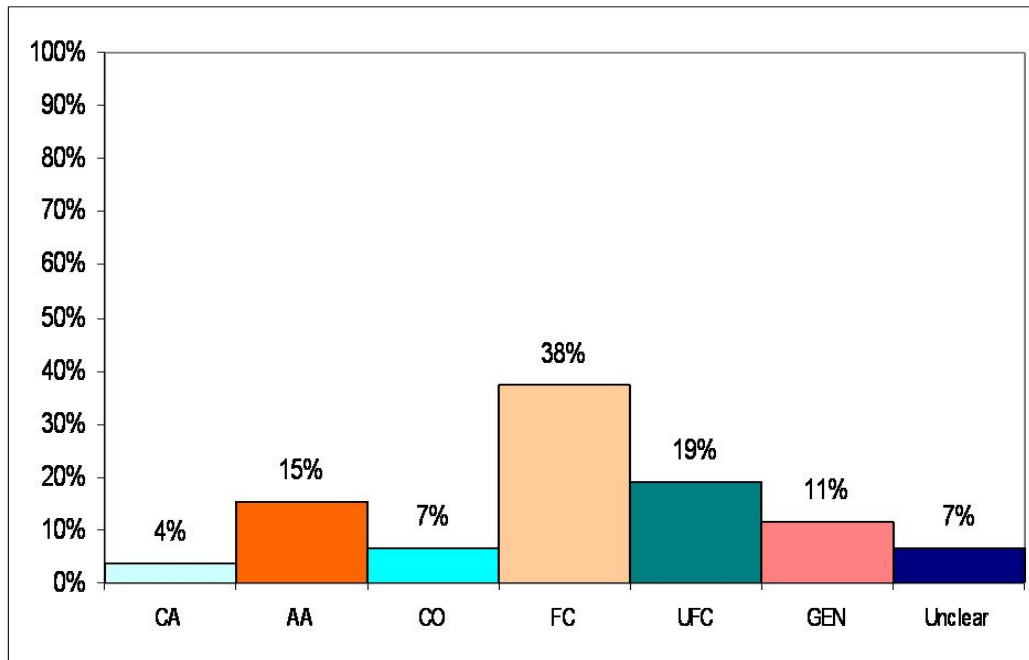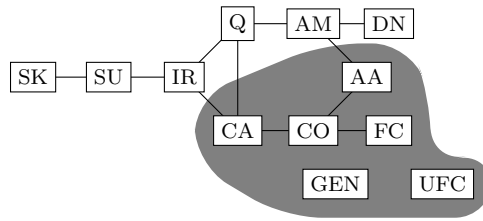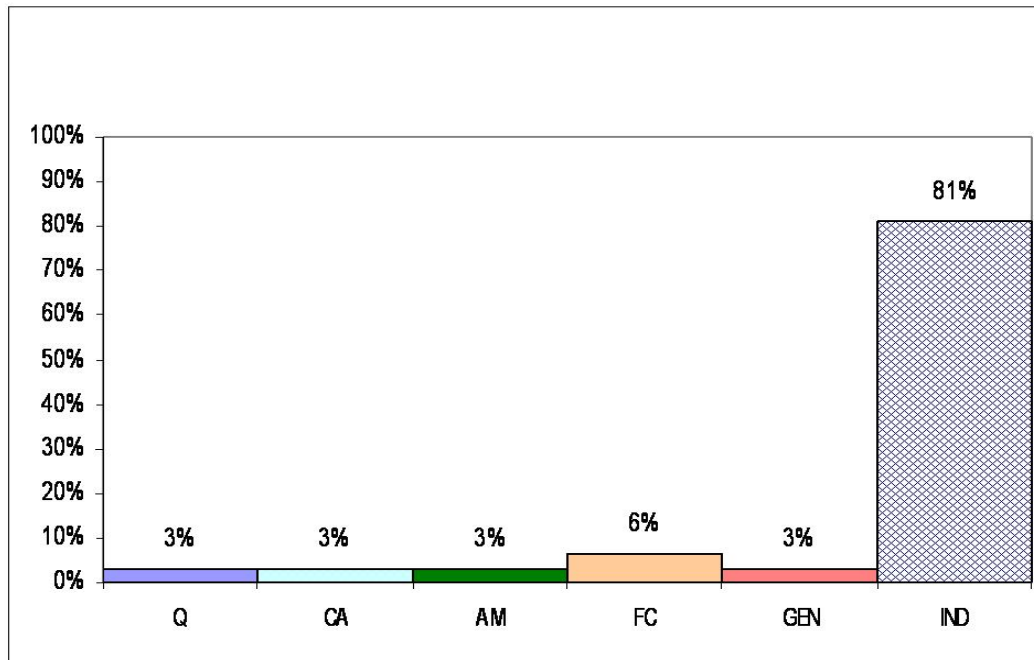

Uno + qualunque + N

(29)   *Distribution*
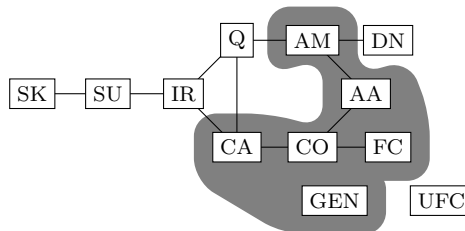
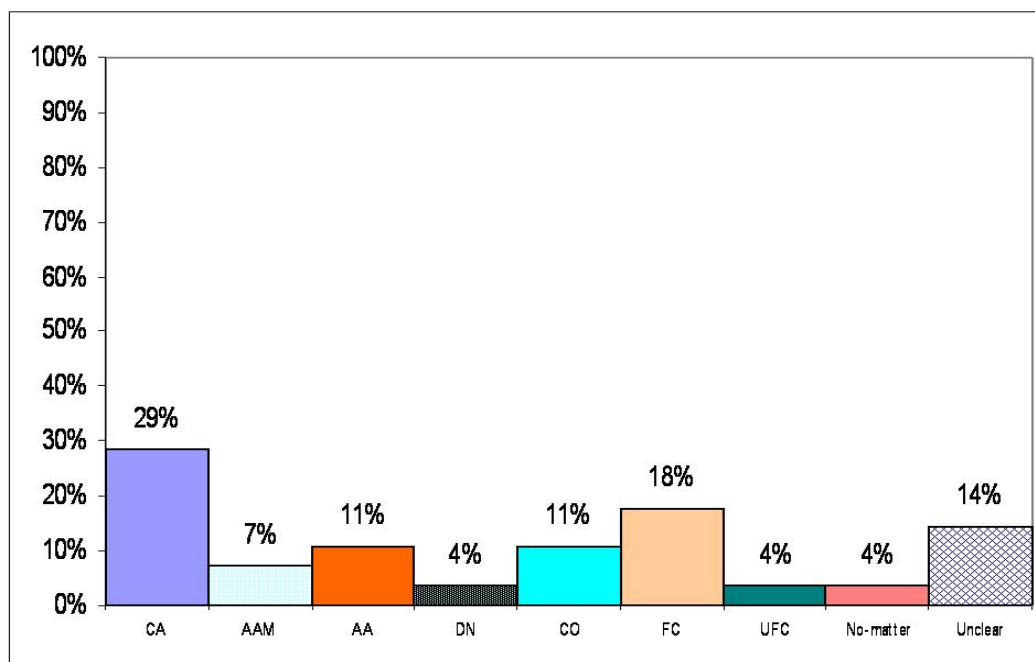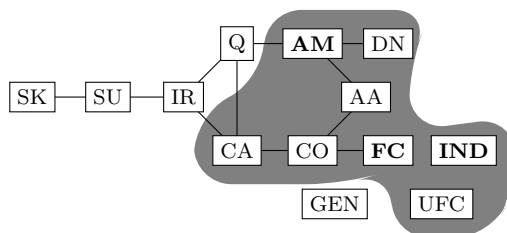**Det + N + qualunque**

(30) *Distribution*

Dutch

- Item: *wie dan ook* [*wie* 'who' + *dan* 'then' + *ook* 'also']
- Corpus: Corpus Gesproken Nederlands CGN (Spoken Dutch Corpus; 10 million words)
- Query: wie dan ook
- Number of occurences: 29

- Labeled: all 29 occurrences

(31)   *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.

4   Diachronic research

In this section we present our findings on the historical development of Spanish *cualquiera* and Dutch *wie dan ook*, two constructions that share the property of employing wh-morphology to express free choice meanings.

4.1   Spanish 'cualquiera'

The Spanish diachronic study consisted of the analysis of occurrences of *cualquiera* (pronoun), or *cualquier* (determiner) in the Spanish historical corpus *El Corpus del Español* created by Mark Davies. We randomly selected 100 occurrences of the item from four periods, 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), translated to English as *whatever*, *whichever*, *whoever* or *any*, and composed of *cual* ('which/who') plus *quier(a)* ('want:3.PRES.SUBJ'), 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).

(32)   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

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. The following graph shows per million of words the number of occurrences of the construction in *El Corpus del Español* in four representative centuries of the history of the language.

(33)   Number of occurrences of 'cualquiera' per million of words[4]
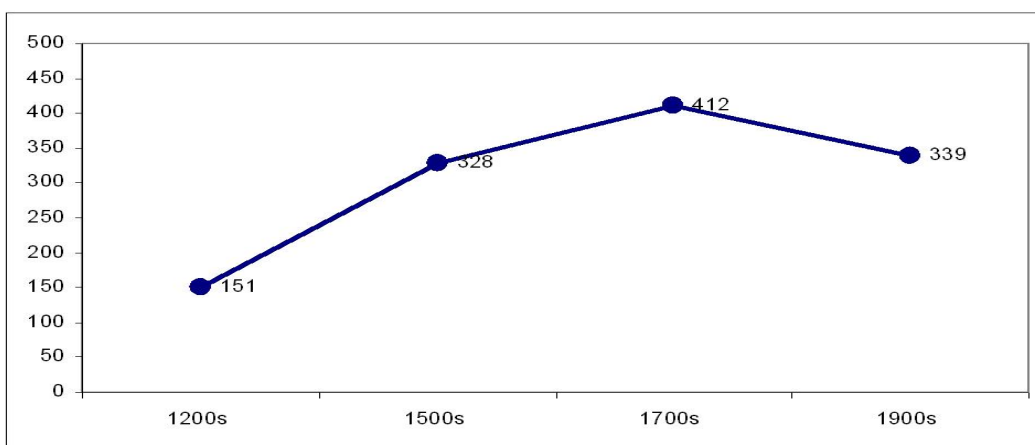
As it can be seen, the presence of *cualquiera* duplicates between the 1200s and the 1500s and reaches a similar proportion to that documented for the 1900s. We take this as an indicator that the use of the construction is consolidated at least since the sixteenth century. The distribution of the functions that *cualquiera* covers throughout these periods points out to a similar conclusion:

(34)   Functions covered by 'cualquiera' in 1200s, 1500s, 1700s and 1900s
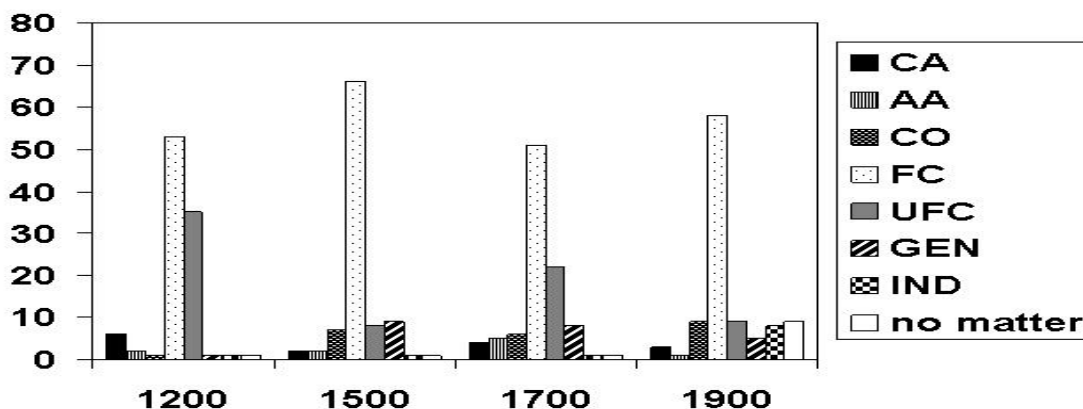

The most noteworthy observation about this distribution is that, generally speaking, it has remained pretty similar throughout the four periods. The FC function is clearly the most dominant since the first period, but some other functions contiguous in the map, namely, CA, CO and AA, as well as the functions 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

---

[4]The actual number of occurrences is 1012 for the 1200s, 5591 for the 1500s, 4048 for the 1700s, and 7744 for the 1900s.

grammaticalization, occurs less and less frequently accompanied by post-nominal modifiers such as restrictive relative clauses and prepositional adjuncts, which typically serve as licensor of free choice items in UFC uses (e.g. *John kissed any woman #(with red hair)*). The graph in (35) reveals this tendency for the sample studied.

(35)   Frequency of post-nominal modification
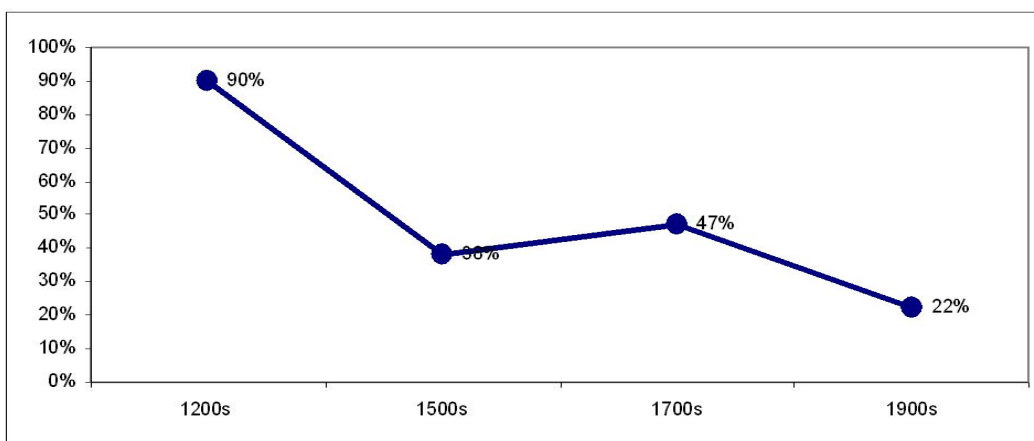
One last important observation is that two more off-map functions, namely IND and *no-matter*, appear in the 1500s and gain presence by the 1900s. The late emergence of the *no-matter* function will turn particularly interesting in light of the development of the Dutch indefinite *wie dan ook*.

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*.

4.2   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:

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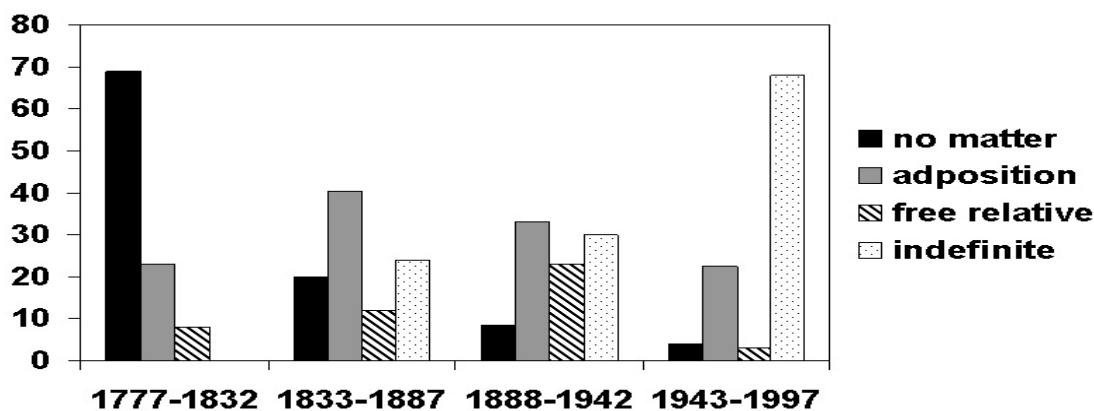
Stage I   The first phase in the grammaticalization of *wie dan ook* as an indefinite is formed by three forms of the *no matter*-function. Characteristic of types of *no matter* constructions is that the *wh dan ook* is not part of the main clause yet: they all consist of either a wh-clause and a main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

(37)   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 (36)).

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.

(38)   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 (36) 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 (39):

(39)   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*]...].

(40)   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 ((36)) illustrates. Here is the distribution of *wie dan ook* in stage IV:

(41)   Functions covered by 'wie dan ook' in stage IV
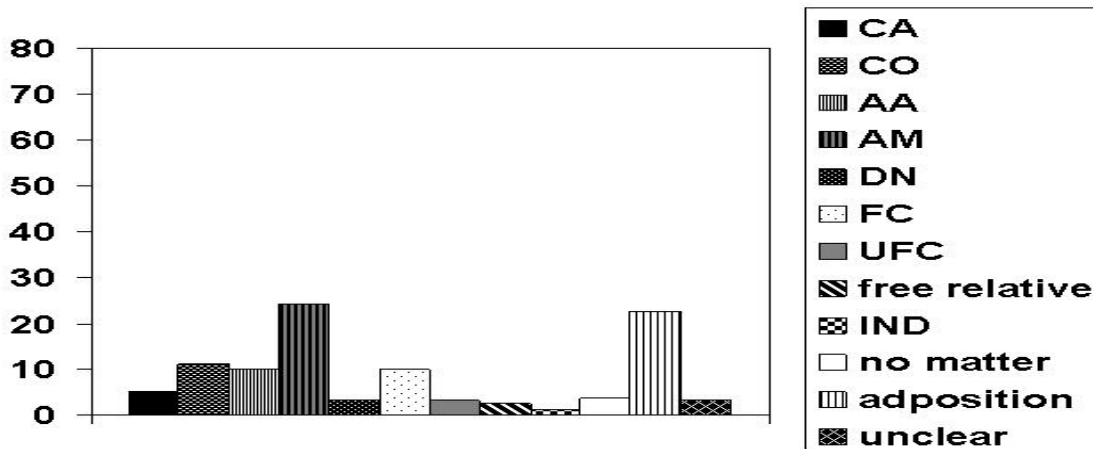
Summarizing: Overall, what can be concluded is that the process of grammaticalization of *wie dan ook* as an indefinite roughly followed four stages, starting off as a *no-matter* construction in a separate wh-clause, slowly evolving into an adpositional modifier on its own, while also turning into a 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're 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).

(42) 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)

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.

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. An interesting conclusion of the diachronic research was that the acquisition of new functions is

---

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

not unidirectional. These studies could not confirm, but neither reject, our initial hypothesis on implicature fossilization. In this article we focused on the battery of tests that we have used during the annotation.

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 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.

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.

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 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 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.

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.