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 and Hedde Zeijlstra

Beyond Semantics
DGfS Workshop

Göttingen
February 23-25, 2011

Corpus studies on indefinites: Motivation

- **Formal pragmatics:** Use of plain indefinites (e.g. *somebody*) can give rise to different pragmatic effects:
  - *Free choice implicature*: each individual is a permissible option (E.g. 'You may invite somebody')
  - *Ignorance implicature*: speaker doesn't know who (E.g. 'Somebody called')
  - ...
- **Typology:** Many languages have developed specialized forms for such enriched meanings:
  - *Free choice indefinites*: Italian *-unque*-series, Czech *koli*-series,
  - *Epistemic indefinites*: Russian *to*-series, German *irgend*-series,
  - ...
- Main hypothesis: Different indefinites as conventionalization (or fossilization) of different pragmatic effects

Illustration main hypothesis: epistemic indefinites

(1) *Plain indefinite (German)*

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

b. Conventional meaning: Someone called

c. Ignorance implicature: The speaker does not know who

(2) *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 epistemic indefinites, inference (1-c), pragmatic in origin, integrated into the semantic content of sentences like (2-a).

Illustration main hypothesis: free choice indefinites

(3) *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

(4) *Free choice determiner (Spanish 'cualquier')*

a. Puedes traer **cualquier** libro.
can:2SG bring:INF any book

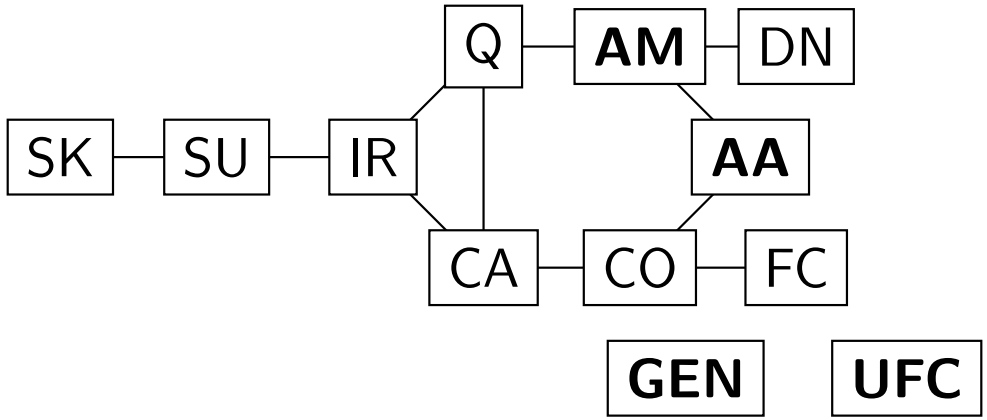
b. Conventional meaning: You can bring me a book and each book is a possible option

In languages with distinctive Free Choice forms, inference (3-c) pragmatic in origin, integrated into the semantic content of sentences like (4-a).

Corpus study on indefinites

- Main objective: Full understanding of
  - what is fossilized (synchronic)
  - how it happened (diachronic)
- Indefinite forms:
  - German EI *irgendein* (synchronic)
  - Czech FC *kterýkoli*
  - Italian FC *(uno) qualunque*
  - Spanish FC *cualquiera*
  - Dutch FC *wie dan ook*
  - Spanish FC *cualquiera* (diachronic)
  - Dutch FC *wie dan ook*
- Methodology
  - 5 coders annotated randomly selected occurrences of the indefinite according to a number of categories
  - Starting point: Haspelmath's functional map

An extended version of Haspelmath's 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.
c.	IR	irrealis	You must try *somewhere* else.
d.	Q	question	Did *anybody* tell you anything about it?
e.	CA	conditional antec.	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.

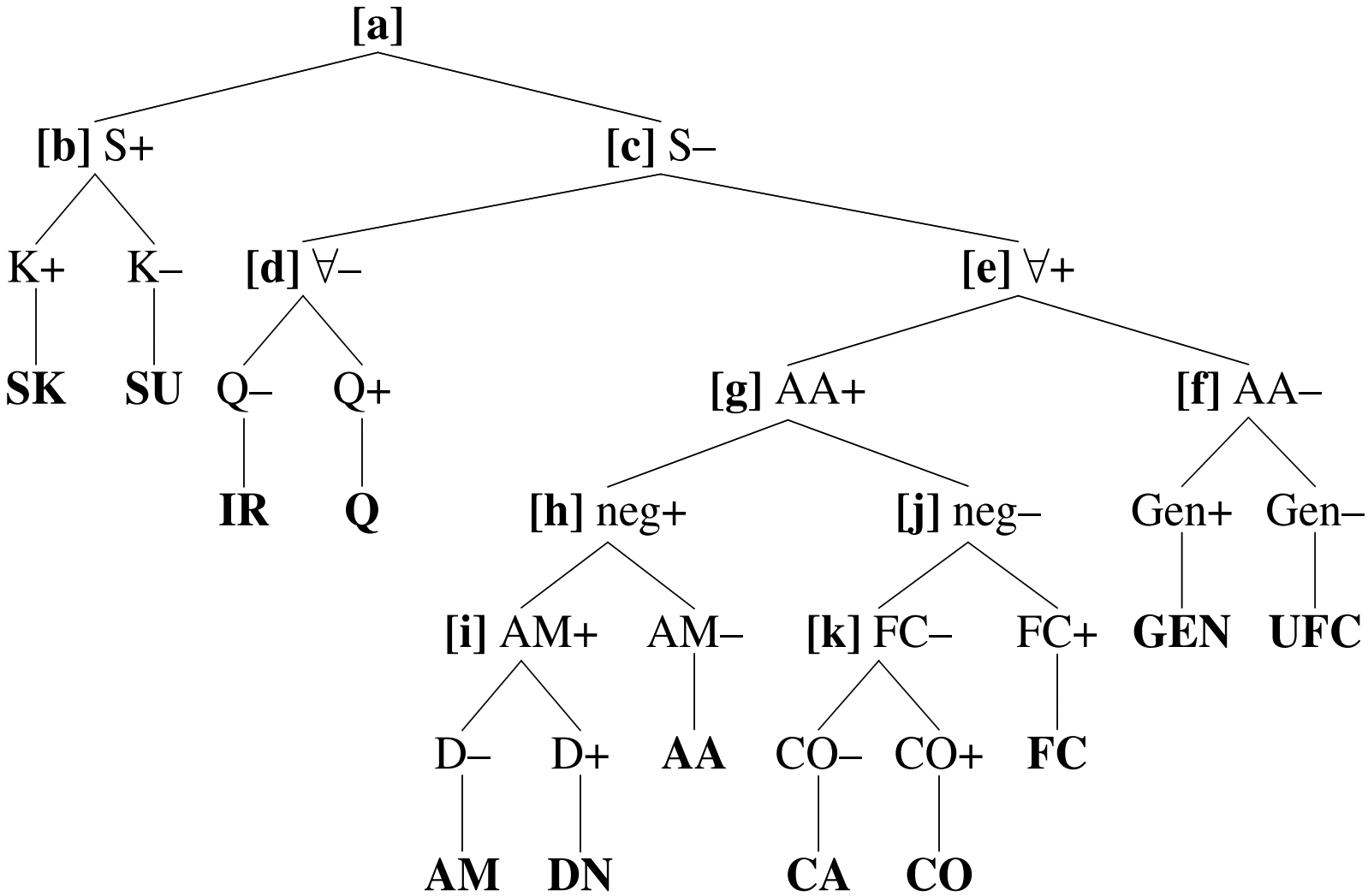
Methodology

- In order for an indefinite to qualify for a function, it must
  - be grammatical in the context the function specifies. E.g. no SK/SU for *any*:

    (5) Somebody /# anybody called. [SK/SU]
  - have the meaning that the function specifies. E.g. no CO for *some*:

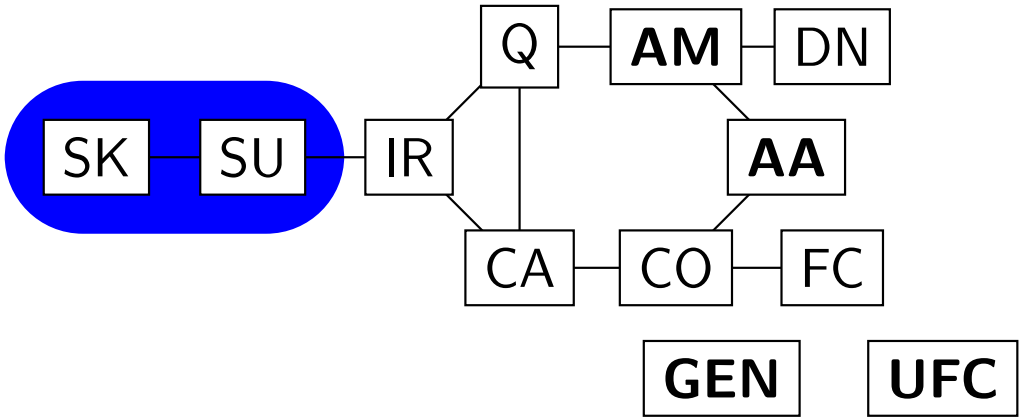
    (6) Berlin is bigger than any /# some Czech city. [CO]
    'For all Czech cities it holds that Berlin is bigger than they are.'
- Extended Haspelmath's functions identified with logico-semantic interpretations
- Diagnostic tests used during annotation organized in a decision tree

Decision tree



Specific–non specific: test [a]

- Specificity area:



- Continuation test [a]: (. . . indefinite$_i$ . . .). (. . . pronoun$_i$ . . .)

(7) SK/SU: I heard *something*. It was very loud. [specific]

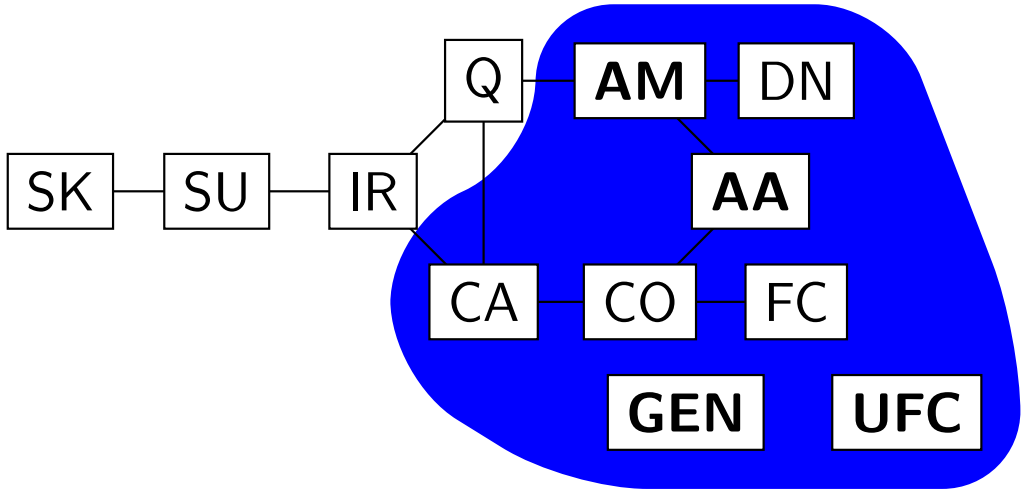
(8) IR: You must try *something* else. # It is very nice. [non specific]

- Standard Analysis:

(9) a. Specific uses: wide scope existential
b. Non-specific uses: narrow scope existential

Existential–wide scope universal: test [c]

- Wide scope universal area:



- Test [c]: **Op** (. . . indefinite . . .) $\Rightarrow \forall x$ (**Op**. . . x . . .)

(10) IR: You must try *somewhere* else $\not\Rightarrow$ for every place x: you must try x [NO]

(11) Q: Did *anybody* tell you anything about it? $\not\Rightarrow$ for every x: did x tell you about it? [NO]

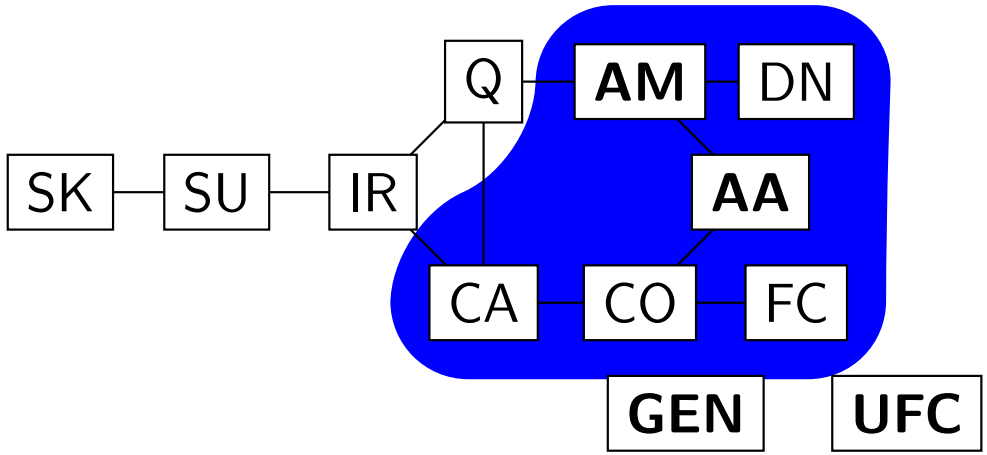
(12) DN: I didn't see *anybody* $\Rightarrow$ for every x: I didn't see x [YES]

(13) FC: You may kiss *anybody* $\Rightarrow$ for every x: you may kiss x [YES]

(14) CA: If you see *anybody*, tell me immediately $\Rightarrow$ for every x: if you see x, tell me immed. [YES]

Anti-additivity: test [e]

- Anti-additive area:



- Anti-additivity test [e]: $Op(a \vee b) \Rightarrow Op(a) \wedge Op(b)$

(15) FC: You may kiss John or Mary $\Rightarrow$ you may kiss John and you may kiss Mary [YES, but not in classical modal logic]

(16) UFC: [John kissed *any* woman with red hair] John kissed Lee or Bea $\not\Rightarrow$ John kissed Lee and John kissed Bea [NO]

(17) DN: I didn't see John or Mary. $\Rightarrow$ I didn't see John and I didn't see Mary [YES]

(18) CO: Bill is taller than John or Mary. $\Rightarrow$ Bill is taller than John and Bill is taller than Mary [YES]

- Within anti-additive area we can distinguish:
  - Negative area (blue): $Op(a \vee \neg a)$ is $\bot$ (test [g])
  - Restrictor area (red): $Op(a \vee \neg a)$ is $\top$
  - Free choice area (yellow): $Op(a \vee \neg a)$ is neither (test [j])

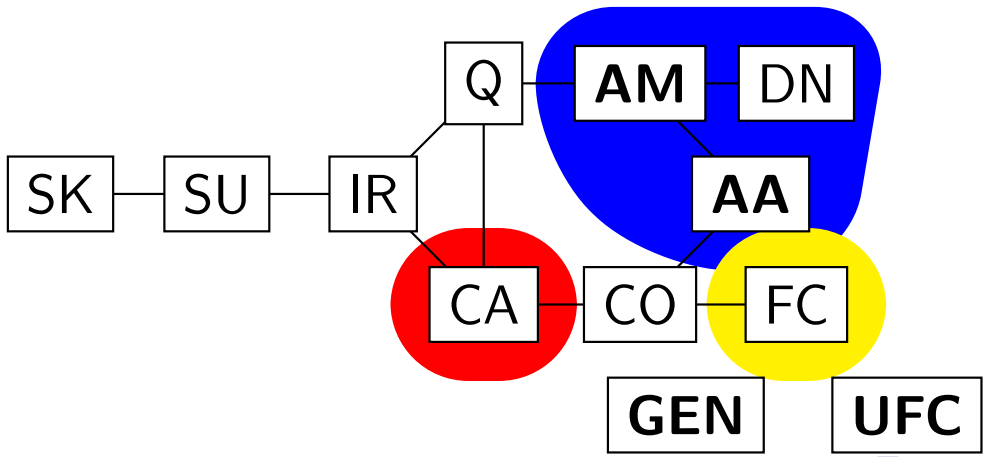
(19) DN: The door is not open or close. (inconsistent)

(20) IN: It is not necessary that (the door is open or close) (inconsistent)

(21) CA: If the door is open or close, I will go to the party. (antecedent is trivial)

(22) FC: The door may be open or close. (informative)

(23) CO: ?Drinking is better than smoking or non-smoking.



Assessment methodology (kappa scores)

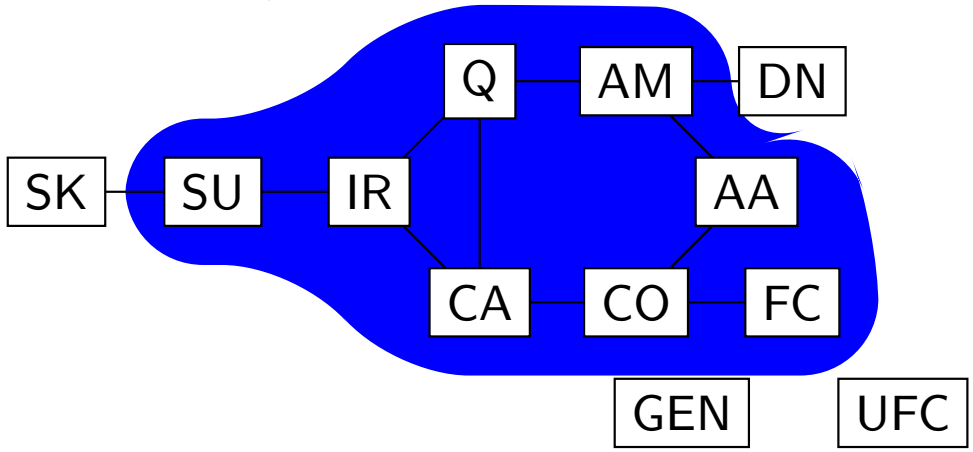
- 5 annotators coded 100 randomly chosen examples from British National Corpus (BYU-BNC): 80 for *any* + 20 for singular *some*
- Annotation was done in three batches (25+25+50) in Jan 2011
- Kappa scores for the different batches of annotation (no weighting)

Items	Kappa
First 25	0.54 (std dev=0.096)
Second 25	0.59 (std dev=0.104)
Last 50	0.46 (std dev=0.087)
Combined 100	**0.52** (std dev=0.069)

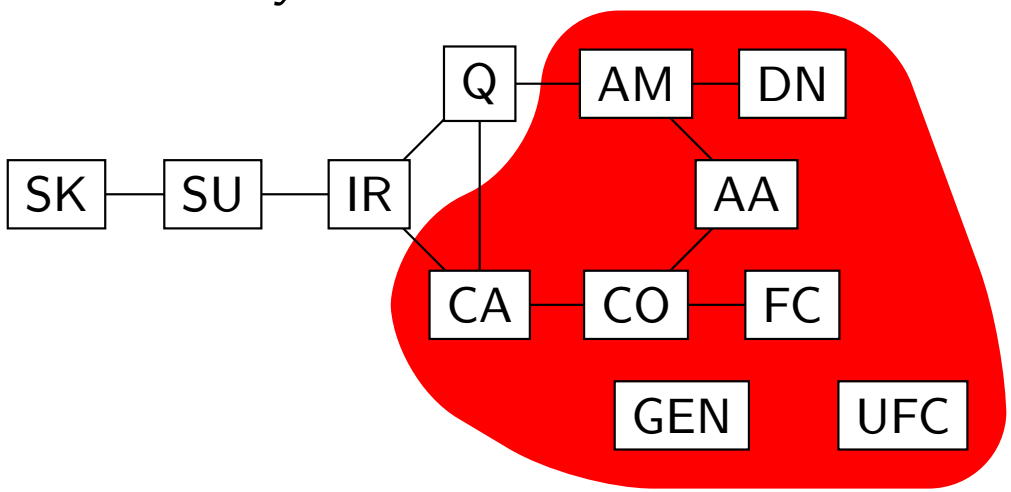
- Kappa score with weighted disagreements: **0.69** (std dev= 0.106)
  - Disagreements not taken into account (had a weight of 0):
    - among the three negative labels (AM, AA and DN)
    - and among the two specific labels (SK and SU)
  - Disagreements considered half correct (weight of 0.5):
    - between the specific functions and IR

Synchronic study: attested distributions

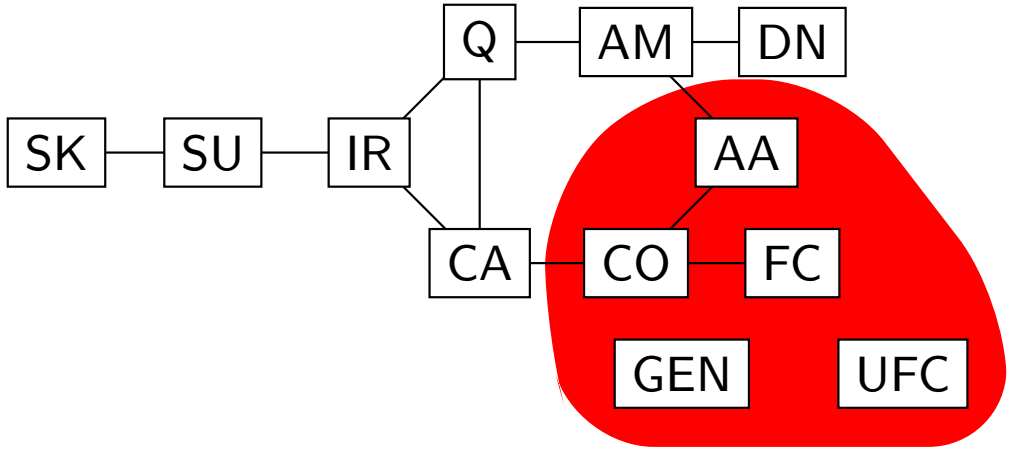
- German *irgendein*



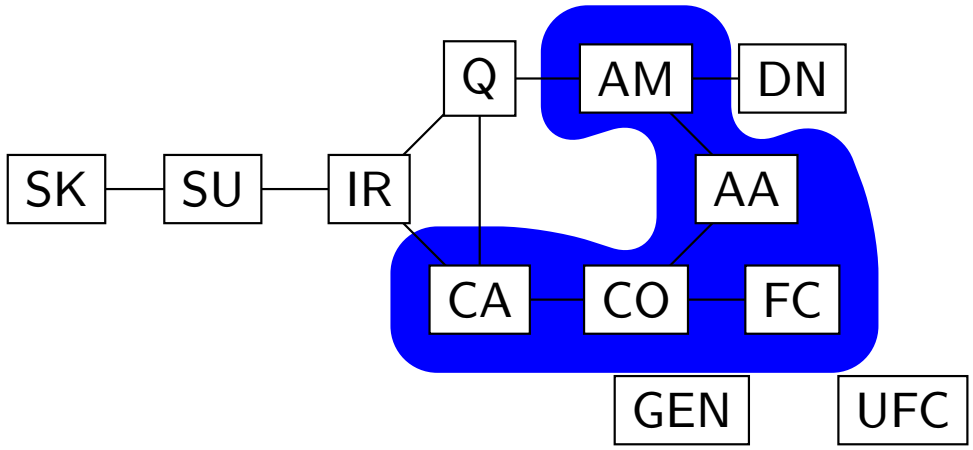
- Czech *kterýkoli*



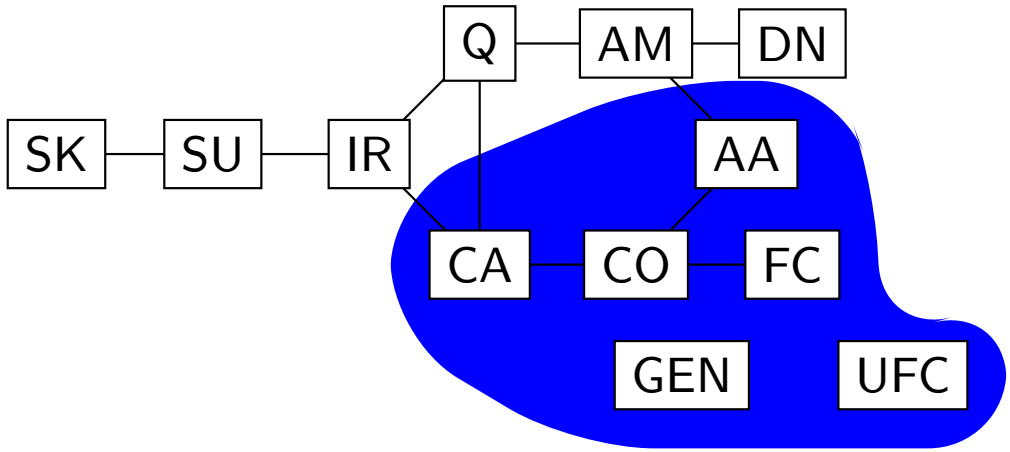
- Italian *qualunque*



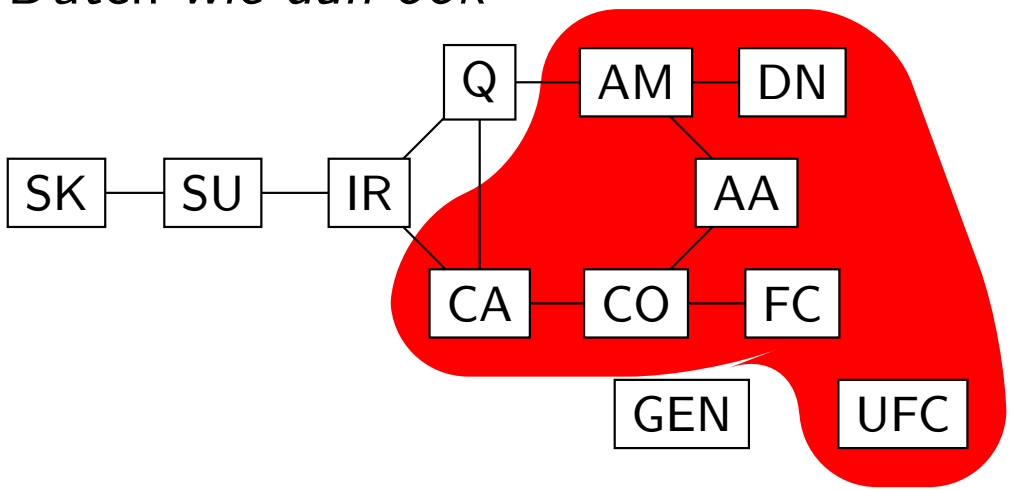
- Italian *uno qualunque*



- Spanish *cualquiera*



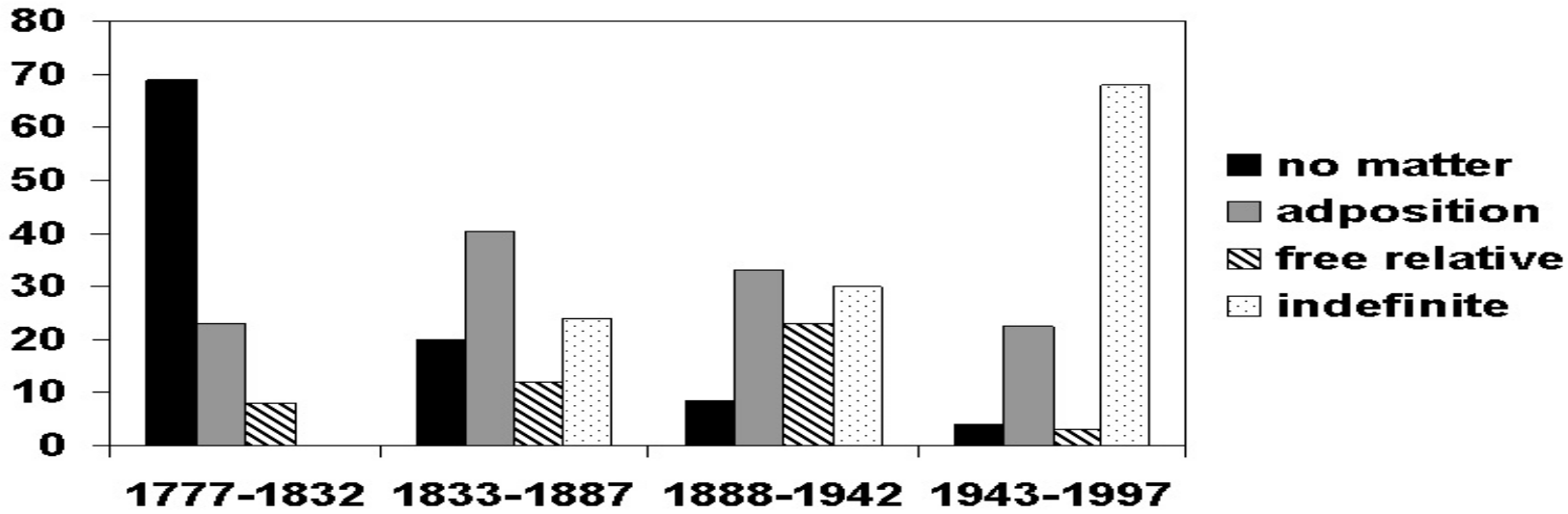
- Dutch *wie dan ook*



Diachronic study: Dutch

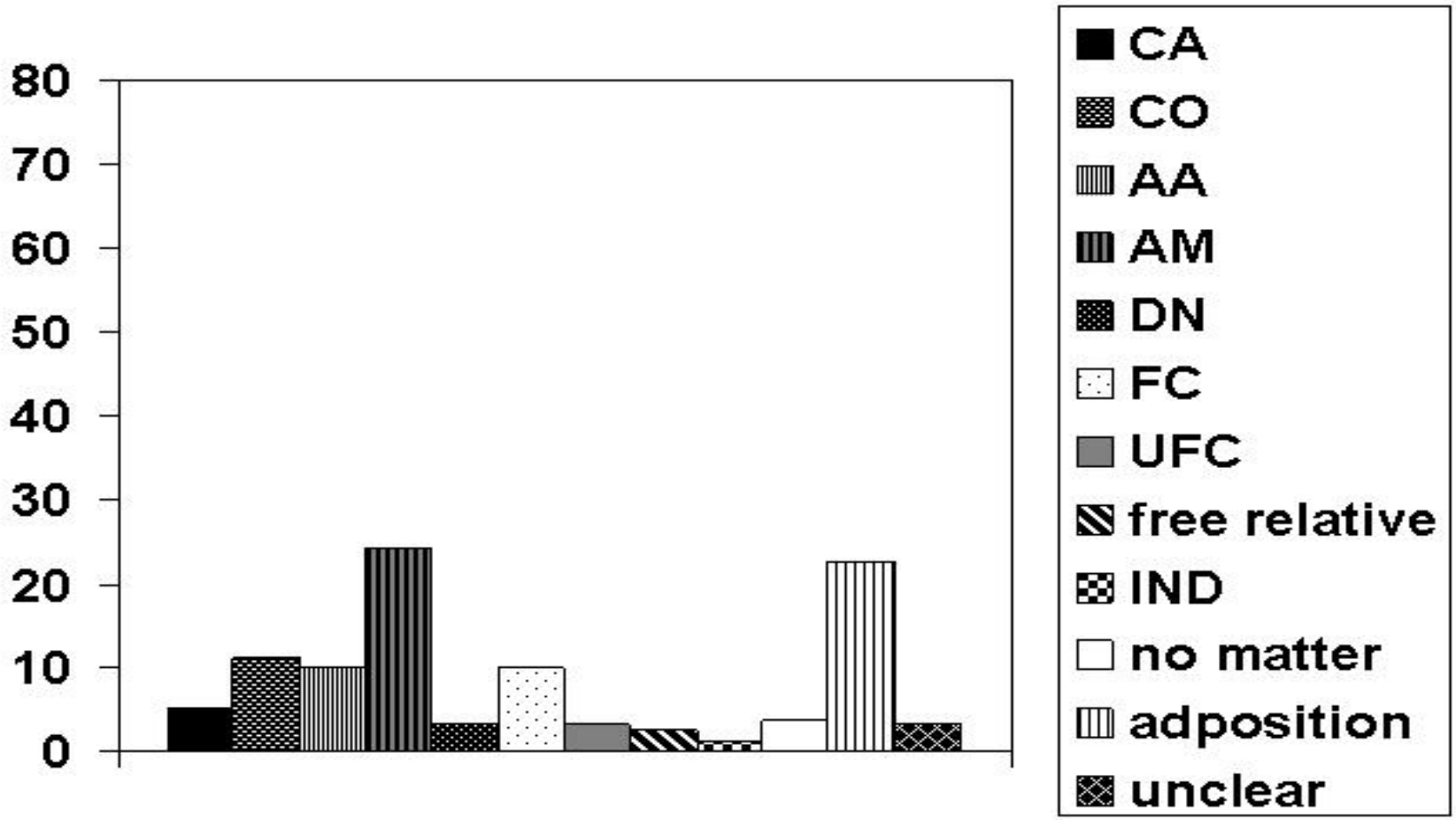
- Item: *wie dan ook* ('who also then')
- Corpus: written Dutch historical corpora
  - CD-ROM Middelnederlands (270 texts before 1300)
  - DBNL (Digitale Bibliotheek voor de Nederlandse Letteren) (4458 texts from 1170-2010)
- Number of occurrences: 349
- Labeled: 349
- The first occurrence found is from 1777

Four stages in grammaticalization of *wie dan ook*



- Stage I: **no matter**
  (24) Wie dan ook naar het feest komt; ik zal blij zijn.
  'Whoever comes to the party; I will be happy.'
- Stage II: **adposition**
  (25) 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.'
- Stage III: **free relative**
  (26) Wie dan ook naar het feest komt, zal blij zijn.
  'Whoever comes to the party(,) will be happy.'
- Stage IV: **indefinite**
  (27) Je mag wie dan ook uitnodigen voor het feest.
  'You may invite anyone to the party.'

Functions covered by ‘wie dan ook’ in stage IV



Discussion

- Initial hypothesis: FC indefinites emerged as the result of a process of conventionalization of an originally pragmatic inference
- Hard to test, not confirmed, but neither rejected
- A possible path consistent with our hypothesis:

  (I) plain indefinite with **conversational** implicature

  (28) Jij mag iemand uitnodigen.

  (II) Plain indefinite + appositive with **conventional** implicature

  (29) Jij mag iemand, wie dan ook (hij mag zijn), uitnodigen.

  (III) New FC indefinite form

  (30) Jij mag wie dan ook uitnodigen

Appositive *wie dan ook* as a new form which expresses the original implicature and later gets grammaticalized

Conclusions

- Report on cross-linguistic synchronic and diachronic corpus study on free choice and epistemic indefinites
- Motivating hypothesis: FCI and EI as fossilization of originally pragmatic inferences
- Methodology:
  - Typologically motivated categories: Haspelmath's map
  - Annotators guided by linguistic tests organized in a decision tree
- Main results:
  - Reliability diagnostic tests: poor (kappa: 0.52) in general, but fair (kappa: 0.69) if internal distinctions within the specificity area and the negative area are disregarded
  - Haspelmath's contiguity hypothesis: confirmed by synchronic study
  - Fossilization hypothesis: neither confirmed nor rejected by diachronic study