Indefinites as fossils

Maria Aloni

[Joint work with Ana Aguilar-Guevara, Angelika Port, Radek Simik, Stephanie Solt, Machteld de Vos and Hedde Zeijlstra]

24 February 2016

Introduction

Main goal report on diachronic corpus studies on indefinites with Free Choice (FC) uses. Data available from:

http://maloni.humanities.uva.nl/Indefinites/corpus.html

Outline

- ▶ Motivation
- ▶ Research questions and hypotheses
- ▶ Methodology
- ▶ Diachronic studies
  - ▶ Spanish *cualquier(a)*   (Aguilar-Guevara, UU)
  - ▶ Dutch *wie dan ook*   (de Vos, UvA)
  - ▶ German *irgend*-series   (Port, UvA)
- ▶ Results and discussion
- ▶ Conclusions

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*: Spanish *cualquier(a)*, Dutch *wie dan ook*, Italian *-unque*-series, Czech *koli*-series, . . .
  - *Epistemic indefinites*: German *irgend*-series, Spanish *algun*, Russian *to*-series, . . .
  - . . .
- <u>Attractive idea</u>: Different indefinites as conventionalizations of different pragmatic effects

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

Free Choice

- ▶ Free choice (FC) inferences:
  - (1) a. Disjunction: $\Box/\Diamond(p_1 \vee p_2) \rightsquigarrow \Diamond p_1 \wedge \Diamond p_2$
    - b. Existential: $\Box/\Diamond \exists x \varphi(x) \rightsquigarrow \forall x \Diamond \varphi(x)$

- ▶ Classical examples
  - (2) Deontic FC   [Kamp 1973]
    - a. You may go to the beach or to the cinema.
    - b. $\rightsquigarrow$ You may go to the beach and you may go to the cinema.
  - (3) Epistemic FC   [Zimmermann 2000]
    - a. Mr. X might be in Victoria or in Brixton.
    - b. $\rightsquigarrow$ Mr. X might be in Victoria and he might be in Brixton.

- ▶ Long-standing debate on the status of FC inferences:
  - ▶ Conversational implicatures (Schulz, Alonso-Ovalle, . . .)
  - ▶ Semantic entailments (Aloni, Barker, . . .)
  - ▶ Obligatory/Fossilized pragmatic inferences (Chierchia, Fox & Spector, Aloni & Franke)

- ▶ If we bring indefinites into the picture:
  - ▶ A purely pragmatic or a purely semantic approach is untenable
  - ▶ Differences between epistemic and deontic FC

Free Choice in indefinites: Spanish

(4) *Plain indefinite (Spanish)*

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

    b. <u>Conventional meaning</u>: You can bring me a book

    c. <u>Free choice implicature</u>: Each book is a possible option

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

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

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

Pragmatic inference (4-c) integrated into the semantic content of sentences like (5-a) (similarly with Dutch *ein* vs *wh dan ook*).

Ignorance inference in indefinites: German

(6) *Plain indefinite (German)*
   a. **Jemand** hat angerufen.
      somebody has called
   b. Conventional meaning: Someone called
   c. Ignorance implicature: The speaker does not know who

(7) *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

Pragmatic inference (6-c) integrated into the semantic content of sentences like (7-a).

Free choice vs ignorance in indefinites

- **Total vs partial variation**
  - TOTAL VARIATION: $\forall x \diamond \phi$
    all alternatives in the relevant domain qualify as a possible option
  - PARTIAL VARIATION: $\exists x \exists y (x \neq y \wedge \diamond \phi(x) \wedge \diamond \phi(y))$
    more than one (but not necessarily all) alternatives in the relevant domain qualify as a possible option

- **Free choice vs ignorance**
  - FREE CHOICE: total variation under deontic or other modals
  - IGNORANCE: partial variation wrt epistemic alternatives

- **Cross-linguistic variety**

	partial-deo	partial-epi	total-deo	total-epi
algún	yes	yes	no	no
vreun	no	yes	no	no
irgendein	?	yes	yes	?
cualquier	no	no	yes	yes
wh- dan ook	no	no	yes	yes

Deontic vs epistemic inferences: German

▶ Total variation (FC) under deontic modals:

(8)   a.   Mary musste **irgendeinen** Artz   heiraten.
          Mary had-to IRGEND-ONE doctor marry
      b.   <u>Conventional meaning</u>: Mary had to marry a doctor, any
          doctor was a permitted marriage option for her.

Narrow scope interpretations (forced by stress) of (8-a) incompatible with
situations in which total variation would not hold [Kratzer & Shimoyama
2002, Aloni & Port 2010]

▶ Partial variation (ignorance) under epistemic modals:

(9)   a.   Juan muss in **irgendeinem** Zimmer im   Haus sein.
          Juan must in IRGEND-ONE room   in-the house be
      b.   <u>Conventional meaning</u>: Juan must be in some room of the
          house and the speaker doesn't know which.

Irgendein can be used in situations in which epistemic total variation
would not hold [Aloni & Port 2010, Lauer 2010]

Research questions and hypotheses

- Synchronic picture: (F_0, M_0), (F_1, M_1)
  - $(F_0, M_0) \mapsto$ unmarked form with plain existential meaning giving rise to pragmatic effect
  - $(F_1, M_1) \mapsto$ marked form with enriched meaning (obligatory free choice or ignorance inference)
- Research question: Hoe did (F_1, M_1) emerge?
- Hypotheses: F_1 emerged as result of **grammaticalization** (involving semantic change ('bleaching'), morpho-syntactic reanalysis and phonological reduction)
- At least two options concerning the emergence of M_1:
  1. <u>Lexicalization</u>: a new form with enriched meaning
     - $(F_0, M_0) > (F_0, M_0), (F_1, M_1)$
  2. <u>Semantic change</u>: a new enriched meaning for an old form
     - $(F_0, M_0), (F_1, M_0) > (F_0, M_0), (F_1, M_1)$

     $\mapsto$ **Fossilization**: a pragmatic inference of some expression is being reanalysed by language learners/speakers as part of the lexical semantics of that expression (Traugott & Dasher 2002)

Outlook of our results

- ▶ Results diachronic studies
  - ▶ Grammaticalization: attested for Dutch and German items (conjectured for Spanish)
  - ▶ Fossilization: possibly confirmed only for German deontic FC meaning
- ▶ Evidence for a pluralistic account of modal inferences:
  - ▶ **Free choice inference** derived as
    - ▶ Semantic entailment for Spanish & Dutch (Menendez-Benito)
    - ▶ Fossilized implicature for German (Aloni & Franke)
  - ▶ **Ignorance inference**
    - ▶ result of lexically encoded felicity conditions: pragmatic variation (e.g., Farkas) or CC-shift (Aloni & Port)
- ▶ Implementation in an information-based semantics employing:
  - ▶ Propositional quantifiers $[\forall], [\exists], \ldots$ (Kratzer & Shimoyama)
  - ▶ Implicature calculation and incorporation (Aloni & Franke)
  - ▶ Dynamic epistemic modals (Veltman)
  - ▶ Quantification under conceptual covers (Aloni)

Corpus study on indefinites

- <u>Indefinite forms</u>:
  - German EI *irgend*-series   (synchronic)
  - Czech FC *kterýkoli*
  - Italian FC *(uno) qualunque*
  - Spanish FC *cualquiera*
  - Dutch FC *wh dan ook*
  - English *any* (and *some*)

  - Spanish FC *cualquier(a)*   (**diachronic**)
  - Dutch FC *wie dan ook*
  - German EI *irgend*-series
- <u>Methodology</u>
  - 6 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
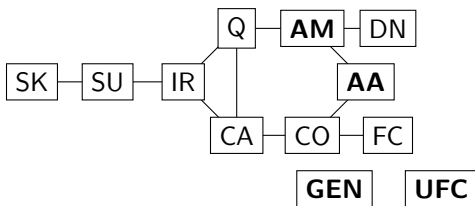
|	Abbr	Label	Example
a.	SK	specific known	*Somebody* called. Guess who?
b.	SU	specific unknown	I heard *something*, but I couldn't tell what.
c.	IR	irrealis	You must try *somewhere* else.
d.	Q	question	Did *anybody* tell you anything about it?
e.	CA	conditional antec.	If you see *anybody*, tell me immediately.
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 SU for *any*:
    - (10) I heard something $/\#$ anything, but I couldn't tell what.   [SU]
  - have the meaning that the function specifies. E.g. no FC for *some*:
   - (11) You may kiss anybody $/\#$ somebody.   [FC]
          'For each individual x it holds that you may kiss x.'
- Extended Haspelmath's functions identified with logico-semantic interpretations
- Diagnostic tests used during annotation organized in a decision tree
- 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   (Aloni et al, LREC, 2012)
- For diachronic studies off-map functions were added: IND, no matter, adposition, free relative,. . .

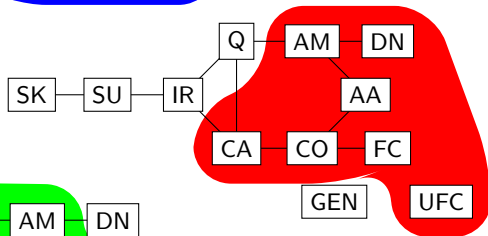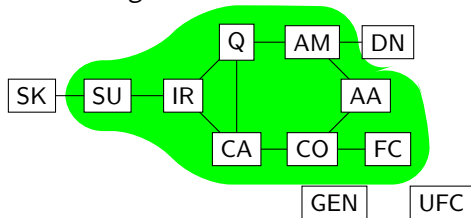# Decision tree
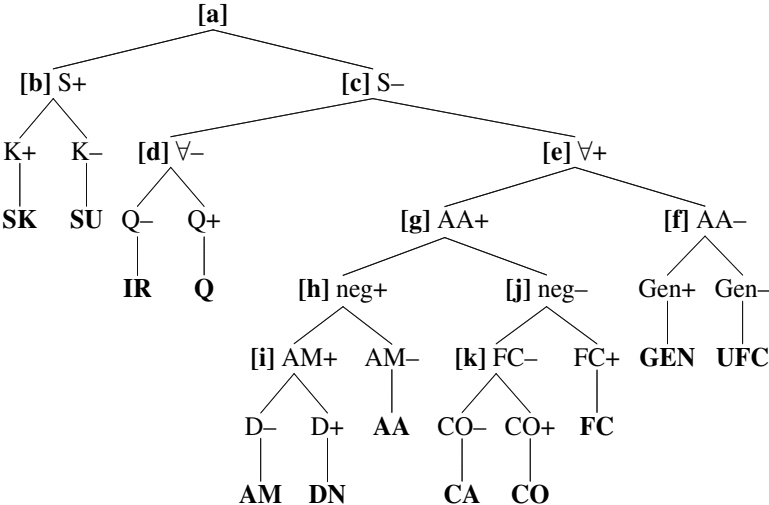
Synchronic study: attested distributions

- Spanish *cualquier(a)*
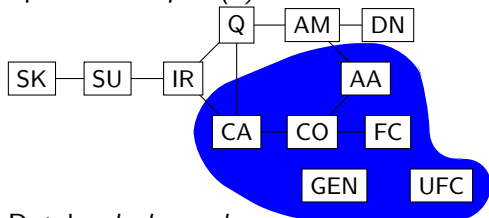
- Dutch *wh dan ook*

- German *irgend*-series

Diachronic study: Spanish (Aguilar-Guevara, UU)

- Item: *Cualquiera* (pronoun), or *cualquier* (determiner), translated to English as *whatever*, *whichever*, *whoever* or *any*, and composed of *cual* ('which/who') plus *quier(a)* ('want:$3.\text{PRES.SUBJ}$')
- Corpus: Spanish historical corpus *El Corpus del Español* created by Mark Davies
- Query: *ualq*
- Occurrences:
  - 1012 for the 1200s (7.9 millions of words)
  - 5591 for the 1500s (19.7 millions of words)
  - 4048 for the 1700s (11.5 millions of words)
  - 7744 for the 1900s (22.8 millions of words)

  These are the four periods in which the history of Spanish has traditionally be divided (Lapesa 1964)
- Labeled: 100 occurrences for each of the first 3 periods, 200 for 1900s

Number of occurrences of 'cualquiera' per million of words
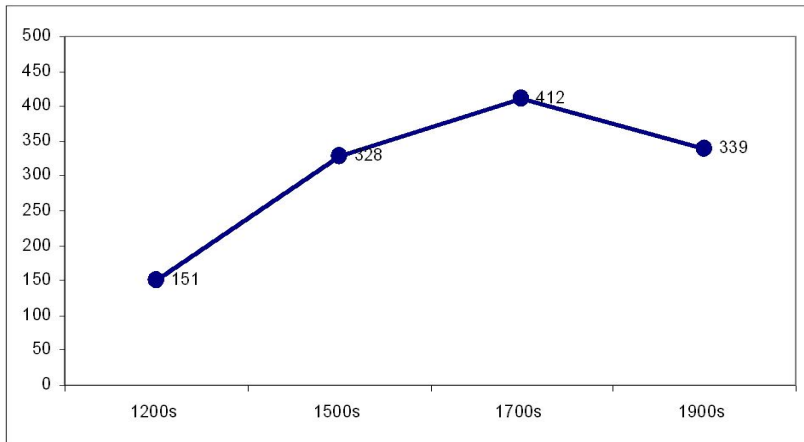
- *Cualquiera*, as a word, already recurrently found in the first documentations of Spanish
- Grammaticalization process could not be attested

Hypothesized grammaticalization process for *cualquiera* (Company-Company and Pozas-Loyo 2009)

(12)   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

Functions covered by 'cualquiera' in four periods
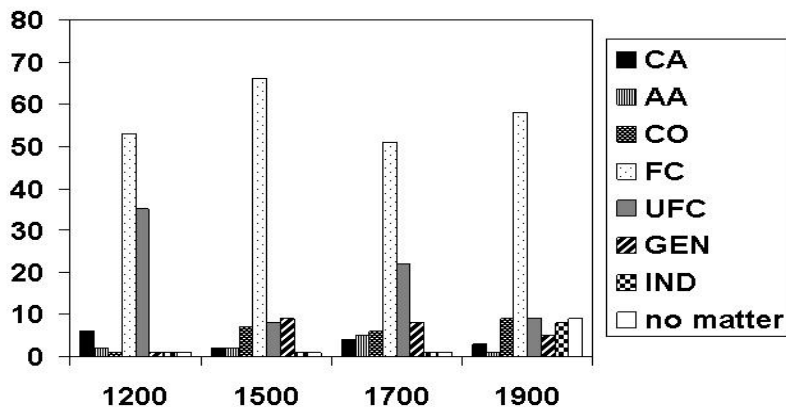
- ▶ The FC function is clearly the most dominant since the first period
- ▶ Two more off-map functions, namely IND and *no-matter*, appear in the 1500s and gain presence by the 1900s
- ▶ The UFC function displays a remarkable decrease starting in the 1500s

Conclusion on Spanish

- Main results
    - *Cualquiera*, as a word, already recurrently found in the first documentations of Spanish
    - Distribution of *cualquiera* stable throughout the four periods (free choice function prominent throughout)
- Given the early grammaticalization and stable distribution, we could not really attest much of the process of *cualquiera* went through in order to behave as it does nowadays
- Conjecture
    - emergence of free choice as result of **lexicalization**, not of semantic change (fossilization)

Diachronic study: Dutch (Machteld de Vos, UvA)

- 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

The first occurrence in 1777

(13)   Het gevoelen dat de Demons, of de Zielen der overleden
       menschen, zulks zouden uitwerken, of dat het, **wie dan ook** de
       Demons der Ouden waren, aen ene bovennatuurlyke oorzaek zou
       toe te schryven zyn, gaet de Autheur hier ten sterkste tegen,
       door ene redenering, die te gelyk ten klaerste toont, dat men
       hier genoegzamen grond heeft, om in natuurlyke oorzaken te
       berusten; zonder dat de Rede ons enigzins verplicht, om op
       bovennatuurlyke oorzaken te denken.   [label: *no-matter*]

[source: Vaderlandsche Letteroefeningen, p. 383; year: 1777]

Four stages in grammaticalization of *wie dan ook*

1. **no matter**   **[wie dan ook + predicate], [main clause]**

   (14)   Wie dan ook naar het feest komt; ik zal blij zijn.
          'Whoever comes to the party; I will be happy.'

2. **adposition**   **[. . . , [wie dan ook], . . .]**

   (15)   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.'

3. **free relative**   **[[wie dan ook + predicate] (,) VP]**

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

4. **indefinite**   **[. . . [wie dan ook] . . .]**

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

Four stages in grammaticalization of *wie dan ook*

- no matter
- adposition
- free relative
- indefinite

▶ Free choice uses prominent in early phases: 47% of indefinite uses in phase II (1833-1887)

Functions covered by 'wie dan ook' in stage IV
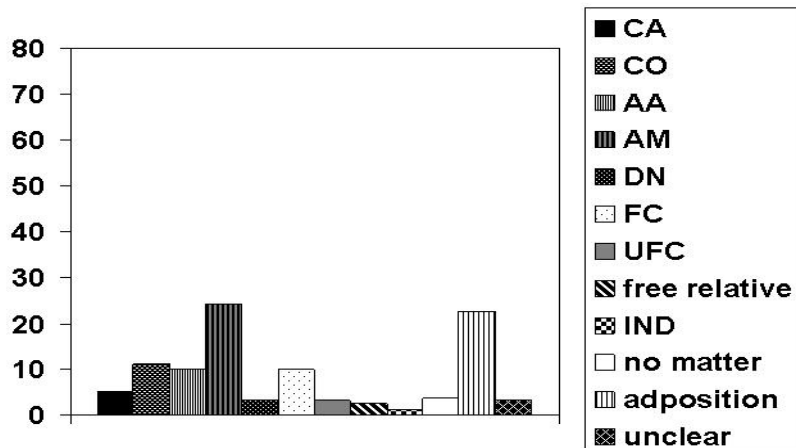
▶ Both FC and NPI uses prominent nowadays

Conclusion on Dutch

- ▶ Main results:
  - ▶ Grammaticalization process involved four subsequent stages, involving semantic, syntactic and phonological change:
    - ▶ no matter > adposition > free relative > indefinite
  - ▶ Free choice uses already prominent in early phases
- ▶ Conclusion: emergence of Dutch free choice as result of **lexicalization**, not semantic change (fossilization)

Towards an analysis of Dutch and Spanish free choice

- ▶ Diachronic data provide evidence for an alternative-based analysis of *wh*-based FC items (Menendez-Benito 2005):
  - ▶ FC items induce propositional alternatives, and require the obligatory application of two covert operators:

    (18) $[\forall] \ldots \mathbf{exh}_{e/st}[\text{FCI}, \lambda x \psi(x)] \ldots$

- ▶ Predictions
  - ▶ FC item correctly predicted to be ungrammatical in episodic sentences and under necessity modals (M-B 2005):

    (19) $\#[\forall](\Box)\mathbf{exh}_{st}[\text{FCI}, \lambda x \psi(x)] \models \bot$

  - ▶ Ready account of FC inferences under possibility modals (derived as entailments):

    (20) $[\forall]\Diamond\mathbf{exh}_{st}[\text{FCI}, \lambda x \psi(x)] \models \forall x \Diamond \psi(x)$

    and also NPI uses, subtrigging (UFC) (Aloni 2007) and universal readings in comparative clauses (CO) (Aloni and Roelofsen 2014)

- ▶ Potential problem: $[\forall]$ and **exh** lacked independent motivation

Discussion: Dutch and Spanish

- In view of diachronic data we can conjecture:
  - **exh** comes from *wh*-morphology
  - emergence of $[\forall]$ triggered by earlier universal-like constructions: no matter or free relative

(21) No matter   (building on Rawlins 2008)

    a.   Wie dan ook naar het feest komt; ik zal blij zijn.

    b. $[\forall]((\mathbf{exh}_{st}[\text{wie dan ook}, \lambda x.\phi(x)])(\lambda_i \,\Box_i\phi))$

        'Whoever comes to the party; I will be happy.'

(22) Free relative   (Aloni 2007)

    a.   Wie dan ook naar het feest komt, zal blij zijn.

    b. $[\forall](P(\downarrow \mathbf{exh}_e[\text{wie dan ook}, \lambda x.\phi(x)))$

        'Whoever comes to the party will be happy.'

- Dutch vs Spanish
  - Dutch: no matter > free relative > indefinite
  - Spanish: free relative > indefinite > no matter

  No unidirectionality!

- (Im)possible developments, if conjecture is correct:
  - # indefinite > no matter, free relatives
  - ? no matter > indefinite > free relatives   (FR phase might be required for syntactic reasons)

Diachronic study: German (Angelika Port, UvA)

- Item: *irgend/irgendein*, in all declinations and spelling variants
- Corpora:
  - Middle High German: Bochumer Middle High German Corpus (BC); Middle High German Conceptual Database (MB)
  - Early New High German: Bonner Early New High German Corpus (BNHG), containing data from 1350-1700; supplemented by other data (corpus of Thomas Gloning (http://www.uni-giessen.de/gloning/etexte.htm), Mediavum (http://www.mediaevum.de/haupt2.htm) and other electronic resources provided by wikisource)
- Query:
  - BC, MB - *irgend* (lemma subsuming different spelling variants, e.g. *irne, jergendt, irgen, yrgend*)
  - BNHG - *rg*. Supplemented by manual searches
- Date of search: 2009
- Number of occurrences:
  - Middle High German – 109 (85 MB, 24 BC)
  - Early New High German – 60 (BNHG 17, other sources 43)
- Labeled: All

The development of *irgend*-indefinites

- Came to life as a locative particle (with restricted distribution) derived from Old High German *io-wergin* glossed as 'somewhere/anywhere'
- Three observed developments:
    - Phase 1: MHG (1050-1350)
      Semantic change (broadening): from locative particle to modal adverb
        - somewhere > sometimes > somehow (1st level of annotation)
    - Phase 2: ENHG (1350-1650)
      Grammaticalization: from particle to non-specific indefinite
        - particle > ambiguous cases > indefinite (2nd level of ann.)
    - Phase 3: NHG (1650-present)
      Semantic change (broadening): establishment of SU and FC functions (3rd level of annotation)

Phase 1: Middle High German (MHG) 1050-1350

- *Irgend* is a particle with a restricted distribution
- Semantic broadening: from locative to temporal to modal
- First ambiguous cases of indefinite uses
- In competition with *n*-particle *nirgend*:

(23)

Function	nirgend	irgend
SK	no	no
SU	no	no
DN	yes	no
IN	yes	no
non-specific	no	yes

- Analysis: *irgend* expresses semantic variation (Farkas)
  - Semantic variation requirement explains #SK, #SU
  - Competition with *nirgend* explains #DN, #IN
- NB: MHG is Negative Concord Language

Phase 2: Early New High German (ENHG) 1350-1650

- First unambiguous occurrences of *irgend*-indefinites
- *Irgendein* enters now into the paradigm of German indefinites together with the plain indefinite *ein* and the negative determiner *kein*

(24)

Function	ein	kein	irgendein
SK	yes	no	no
SU	yes	no	no
DN	no	yes	no
IN	yes	yes	yes
non-specific	yes	no	yes

- Analysis:
  - *irgend* still expresses semantic variation: #SK, #SU
  - Competition with *kein*: #DN
  - IN uses explained by the fact that ENHG in transition from Negative Concorde to Double Negation

Phase 3: Modern High German (MHG) 1650-present

- Negative Concord readings of *kein* no longer possible
- SU and FC uses of *irgend*-indefinites established
  - Emergence of FC can be viewed as result of semantic change triggered by **fossilization** of pragmatic inference
  - Emergence of SU explained in terms of shift from semantic to pragmatic variation

(25)

Function	ein	kein	irgendein
SK	yes	no	no
SU	yes	no	yes
DN	no	yes	no
IN	yes	no	yes
non-specific	yes	no	yes
FC	no	no	yes

- Analysis:
  - *irgend* expresses semantic-pragmatic variation: #SK, SU, IN
  - Competition with *kein*: #DN
  - FC inference derived as fossilized implicature (Aloni & Franke)
    - Fossilized FC inference derived for both possibility and necessity modals ($\neq$ Spanish & Dutch case)
    - But not for epistemic modals (*contra* Chierchia et al)

Conclusions

- ▶ Report on cross-linguistic diachronic corpus study on indefinites with FC uses (exhibiting the FC function)
- ▶ Motivation: shed light on debate on status of (obligatory) modal (FC/ignorance) inferences
- ▶ Research question: Hoe did obligatory modal inference emerge in Spanish, Dutch and German indefinites?
- ▶ Methodology:
  - ▶ Typologically motivated categories: Haspelmath's map
  - ▶ Annotators guided by tests organized in a decision tree
- ▶ Main result:
  - ▶ Fossilization: possibly confirmed only for German deontic FC meaning
- ▶ Evidence for a pluralistic view on modal inferences:
  - ▶ **Free choice inference** derived as
    - ▶ Semantic entailment for Spanish & Dutch (Menendez-Benito)
    - ▶ Fossilized implicature for German (Aloni & Franke)
  - ▶ **Ignorance inference**
    - ▶ result of lexically encoded felicity conditions: pragmatic variation (e.g., Farkas) or CC-shift (Aloni & Port)