


# Classifying conditionals in discourse

CogLing7, 2017

Alex Reuneker, Universiteit Leiden

[a.reuneker@hum.leidenuniv.nl](mailto:a.reuneker@hum.leidenuniv.nl)



IF YOU WANT BREAKFAST  
IN BED TOMORROW,  
SLEEP IN THE KITCHEN  
TONIGHT.

# Outline

- Introduction
- Three classifications of conditionals
- Why study reliability?
- Methodology
- Results
- Conclusions

# Introduction

- Classification: to group phenomena in such a way that within-group variance is minimized and between-group variance is maximized (cf. Bailey 1994)
- Many classifications of conditionals exist (e.g. Quirk et al. 1985; Sweetser 1990; Athanasiadou & Dirven 1997; Declerck & Reed 2001; Dancygier & Sweetser 2005 and more).
- Existing classifications are sometimes criticized for not being applicable to 'real language data' (see e.g. Ferguson 2001; Carter-Thomas & Rowley-Jolivet 2008).

# Three classifications of conditionals

- Three major classifications of conditionals were tested on corpus data:
  - Quirk et al. (in *A Comprehensive Grammar of the English Language* 1985)
    - Direct: *If you put the baby down, she'll scream.*
    - Indirect: *She's far too considerate, if I may say so.*
    - Rhetorical: *If they're Irish, I'm the Pope.*
  - Athanasiadou & Dirven (1996, 1997)
    - Hypothetical: *If there is no water in your radiator, your engine will overheat [...].*
    - Course-of-event: *If the students come on Fridays, they have oral practice [...].*
    - Pragmatic: *There are biscuits on the sideboard if you want them.*
  - Dancygier & Sweetser (2005)
    - Predictive (content): *If it rains, the match will be cancelled.*
    - Epistemic: *If she's divorced, then she has been married.*
    - Speech-act: *There are biscuits on the sideboard if you want them.*
    - Meta-linguistic: *OK, I'll have a tomahto, if that's how you pronounce it.*

# Three classifications of conditionals

| Main types            | Sub types  |                           |
|-----------------------|--|---------------------------|
| Direct condition      | Open condition   |                           |
|                       | Hypothetical condition   | Future<br>Present<br>Past |
| Indirect condition    | Request for permission<br>(Meta-linguistic) comment<br>Extra-linguistic uncertainty<br>Utterance |                           |
| Rhetorical conditions | Projection from main clause<br>Projection from <i>if</i> -clause                                 |                           |

Quirk et al. (1985)

| Main types                  | Sub types                                 |                                 |
|-----------------------------|---|---------------------------------|
| Hypothetical conditional    | Causal                                    |                                 |
|                             | Condition                                 |                                 |
|                             | Supposition                               |                                 |
| Course-of-event conditional | Descriptive<br>Inferential<br>Instructive |                                 |
| Pragmatic conditional       | Conversational                            | Discourse<br>Meta-communicative |
|                             | Logical                                   | Inferencing<br>Identifying      |

Athanasiadou & Dirven (1996; 1997)

| Main types                       | Sub types   |  |
|----------------------------------|---|--|
| Predictive conditional (content) | If-backshift  |  |
|                                  | Hypothetical backshift  |  |
|                                  | Generic   |  |
| Non-predictive conditional       | Epistemic<br>Speech-act<br>Meta-linguistic<br>Meta-metaphorical |  |

Dancygier & Sweetser (2005)

# Why study reliability?

- Reliability (Kaplan & Goldsen 1965): stability, replicability, accuracy

|   | Conditional  | Q                      | A & D            | D & S             |
|---|--|------------------------|------------------|-------------------|
| 1 | If that's art, then I'm an artist too!   | Rhetorical             | Pragmatic        | Epistemic         |
| 2 | If Oscar is selected to study medicine, then a golden future awaits him.   | Direct (open)          | Hypothetical     | Predictive        |
| 3 | If life is a candle, people are the moths flying towards it.   | Indirect (meta-ling.)  | Pragmatic        | Meta-metaphorical |
| 4 | So if I understand correctly, the singer of REM did not leave the band?  | Indirect (extra-ling.) | Pragmatic        | Speech-act        |
| 5 | So: if you're interested and you don't have any plans yet, the Dutch Philharmonic Orchestra plays Tchaikovsky tonight. | Indirect (utterance)   | Pragmatic        | Speech-act        |
| 6 | If there wouldn't be coffee, the guests wouldn't stay this long.   | Direct (hypothetical)  | Hypothetical     | Predictive        |
| 7 | And every time you have worked-out, your energy level has dropped.   | Direct (open)          | Course-of-events | Predictive        |
| 8 | If the O is in the name of the month again, it'll get cold quickly.  | Direct (open)          | Course-of-events | Epistemic         |

# Quality of classifications

- To what extent are classifications *exhaustive* and *mutually exclusive*? (cf. Greenberg 1957; Marradi 1990; Bailey 1994)
- **Exhaustivity:**
  - All conditionals are assigned to a class (i.e. there are no 'left-overs').
- **Mutual exclusivity:**
  - A conditional may be placed in just one class based on a limited set of parameters.
  - 'A given example may be *ambiguous* between interpretations in two different domains, [...], but no *one* interpretation of an *if-then* sentence [...] simultaneously expresses conditionality in more than one domain' (Sweetser 1990: 124-5).
  - Although Athanasiadou and Dirven (1997) and Dancygier (1998) present their classification in terms of prototype theory (cf. Rosch 1978; Lakoff 1987; Taylor 2001), in their results all examples are classed as only one type.



# Why study reliability?

- How to determine which classification is ‘best’?
  - Which classification is right? (*natural* classification, cf. Marradi 1990)
  - Which classification explains the most data most efficiently? (*artificial* classification, cf. Tiryakian 1968; Feger 2001; Van der Auwera & Gast 2013)
- Preliminary for answering any of these questions
  - **To what extent are raters able to apply classifications to (real) data?**
  - ‘Despite its importance, there is presently no tradition in the field of corpus-based discourse studies to report agreement measures’ (Spooren & Degand 2010).
  - The evaluation of a classification is ‘frequently neglected’ (Feger 2001).
  - ‘Objects under study are often found to resist a tidy pigeon-holing of this kind’ (Hempel 1965).
  - The applicability of classifications is challenging, but an important measure of its validity (cf. Bolognesi, Pilgram & Van Den Heerik 2016).
  - Total accountability: taking the responsibility to account for all corpus data present in the corpus or sample, including ambiguous, border-line or unclear exemplars (McEnery & Hardie 2012).

# Methodology: design

- In order to assess to which extent classification results are 'independent of the measuring event, instrument or person' (Kaplan & Goldsen 1965: 83), *reliability* was investigated.
- Experiment:
  - A group of raters classified a set conditionals from
    - the CONDIV-corpus (Grondelaers et al. 2000) of written Dutch
    - the CGN-corpus of Spoken Dutch (Oostdijk 2000)
- The experiment followed an within-subjects design to control for individual differences (e.g. level of linguistic knowledge, understanding of the materials).
- In order to control for memory and practice effects, the experimental conditions were counterbalanced using a latin-square design.
- (So, each subject classified a number of conditionals three times.)

# Methodology: subjects

- The subjects (N=27) were undergraduate students in linguistics following a course on corpus linguistics and conditionals.
- All subjects registered Dutch as their primary language.
- All subjects registered General Linguistics or Dutch Linguistics as major field of study.
- The mean age was 22.67 (SD=5.13).

# Methodology: materials

- The materials consisted of:
  - 3 test items (labeled as such) to familiarize subjects with the task/online presentation
  - 14 items from the written corpus (CONDIV; only Dutch newspaper articles)
  - 9 items from the spoken corpus (CGN; only spontaneous F2F)
  - 8 control items (variations on idealized examples from the literature)
  - (2 re-test items)
- The sentences were selected from a random sample of sentences containing *als* (if).
  - The sample was manually cleared of non-conditional uses of *als* (e.g. comparative use, insubordinate uses).
  - All items included one sentence preceding and one sentence following the conditional sentence.
  - The conditional sentence itself was presented in **bold**.

# Methodology: materials

- Example of written corpus item:
  - Veel mensen doen dingen waarvan ze best weten dat het niet mag. **Maar als het niet wordt bestraft, gaat zoiets wennen en gaat het steeds een stukje verder.** Daar gaan we nu een stokje voor steken.
  - *Many people do things while knowing they're not allowed to. **But if it isn't punished, people will get used to it and the will continue to progress.** That is what we are going to prevent.*
- Example of spoken corpus item:
  - oh. ja ik vind dat wel heel interessant en ik **als ik dat zo zie dan denk ik ook dat die man 't heeft gedaan.** ook als je dat hoort maar van de andere kant ja.  
[fna000458\_\_166]
  - oh. Yeah I find that very interesting and I **if I see it like then I also think that man has done it.** Also if you hear that but on the other side yeah.

# Methodology: procedure

- For each classification, the original texts (Quirk et al. 1985; Athanasiadou & Dirven 1997; Dancygier & Sweetser 2005) were distributed as part of the course materials.
- Each classification was read and every class a set of conditionals (both test and corpus data) was classified and discussed.
- A week before the experiment, students were presented with an overview of the classifications, including criteria and examples for each type.

## INSTRUCTIES DANCYGIER EN SWEETSER (2005)

De bijzin wordt *cursief* weergegeven, de hoofdzin onderstreept.

1. De (waarheid van) de situatie in de hoofdzin is **inhoudelijk-afhankelijk** van die in de *bijzin*.

- 1.1 De werkwoordstijd wordt gebruikt om **onzekerheid** uit te drukken (*backshift*):

Voorbeeld: 'Als Hiro de kaart zal pakken' wordt 'Als Hiro de kaart pakt'  
'Als ik won.' wordt 'Als ik had gewonnen.'

Er wordt een voorspelling gemaakt waarbij **alternatieve situaties** spelen:

Voorbeeld: 'Als je het gras maait, krijg je een beloning.'  
'Als je het gras niet maait, krijg je geen beloning.'

Type: voorspellend

## INSTRUCTIES ATHANASIADOU EN DIRVEN (1997)

De bijzin wordt *cursief* weergegeven, de hoofdzin onderstreept.

1. De (waarheid van) de situatie in de hoofdzin is **inhoudelijk-afhankelijk** van die in de *bijzin* en de situaties zijn **hypothetisch/onzeker**.

De spreker **committeert zich niet** aan de waarheid van de situatie in de *bijzin*.

Er kan (on)waarschijnlijkheid worden aangegeven met o.a. werkwoordstijden.

De zin is te parafaseren met *omdat*, 'op voorwaarde dat' of 'stel dat'.

Voorbeeld: 'Als het goed weer is, gaan we zwemmen.'  
(*'We gaan zwemmen, omdat/op voorwaarde dat het lekker weer is.'*)  
(*'Stel dat het lekker weer is, dan gaan we zwemmen.'*)  
'Als het mooi weer zou zijn, zouden we gaan zwemmen.'

Type: hypothetisch

# Methodology: procedure

- The subjects first rated 3 test items (marked as such) to familiarize them with instructions, task and materials.
- After the test set, each subject classified 23 corpus items and 8 control items (not marked as such) per trial.
- Per trial, all items (corpus and control) were presented in random order.
- Per item, raters chose a class, indicated their certainty on a 5-point scale and optionally included a remark. Raters were able but discouraged to choose 'no answer'.

# Results: selection of qualified raters

- Control items allowed for the selection of students who were able to correctly classify idealized examples.
- Accuracy (i.e. how well a rater's judgments confirm to actual known values) was calculated for each rater by dividing the number of true positives by the total number of classifications made, here  $7/8 = 0.875$  (cf. Stehman 1997).

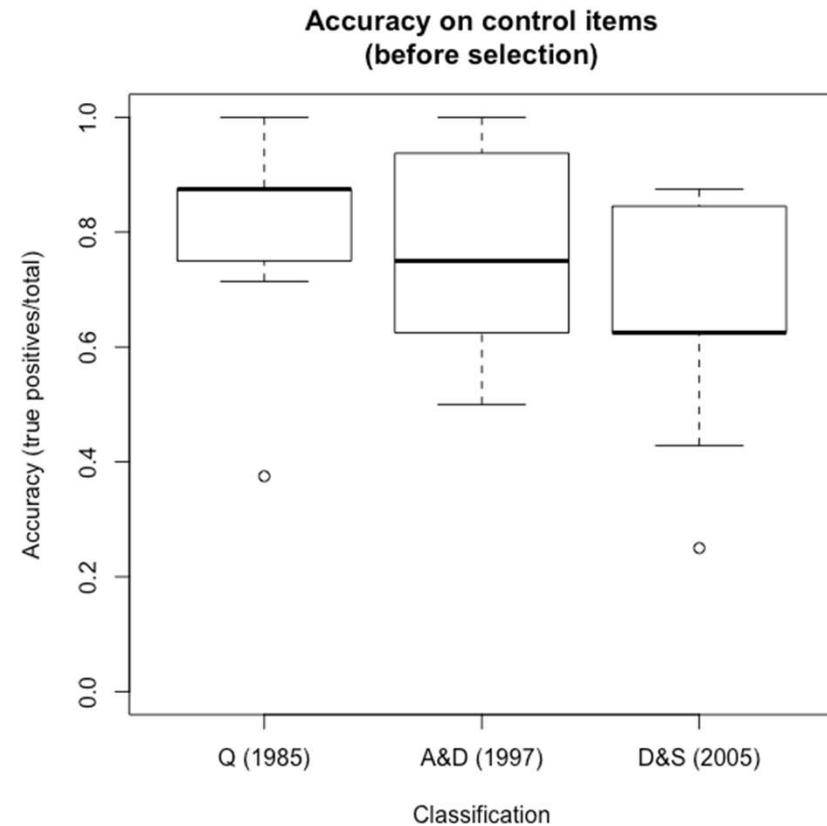
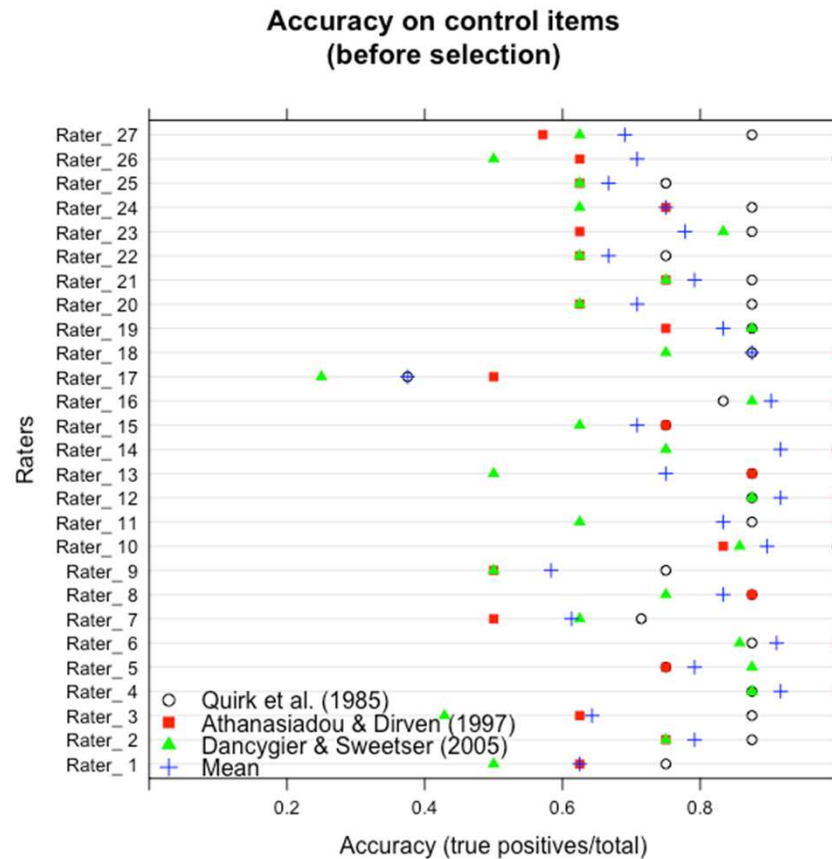
|                      | Hypothetical (R) | Course-of-events (R) | Pragmatic (R) |
|----------------------|------------------|----------------------|---------------|
| Hypothetical (A)     | 2                | 1                    | 0             |
| Course-of-events (A) | 0                | 1                    | 0             |
| Pragmatic (A)        | 0                | 0                    | 4             |

7 correct classifications (true positives), 1 incorrect (misclassification)



# Results: selection of qualified raters

- Distribution of accuracy scores per rater (before selection)



# Results: selection of qualified raters

|                              | Mean accuracy | SD   |
|------------------------------|---------------|------|
| Quirk et al. (1985)          | 0.84          | 0.12 |
| Athanasiadou & Dirven (1997) | 0.81          | 0.17 |
| Dancygier & Sweetser (2005)  | 0.68          | 0.16 |

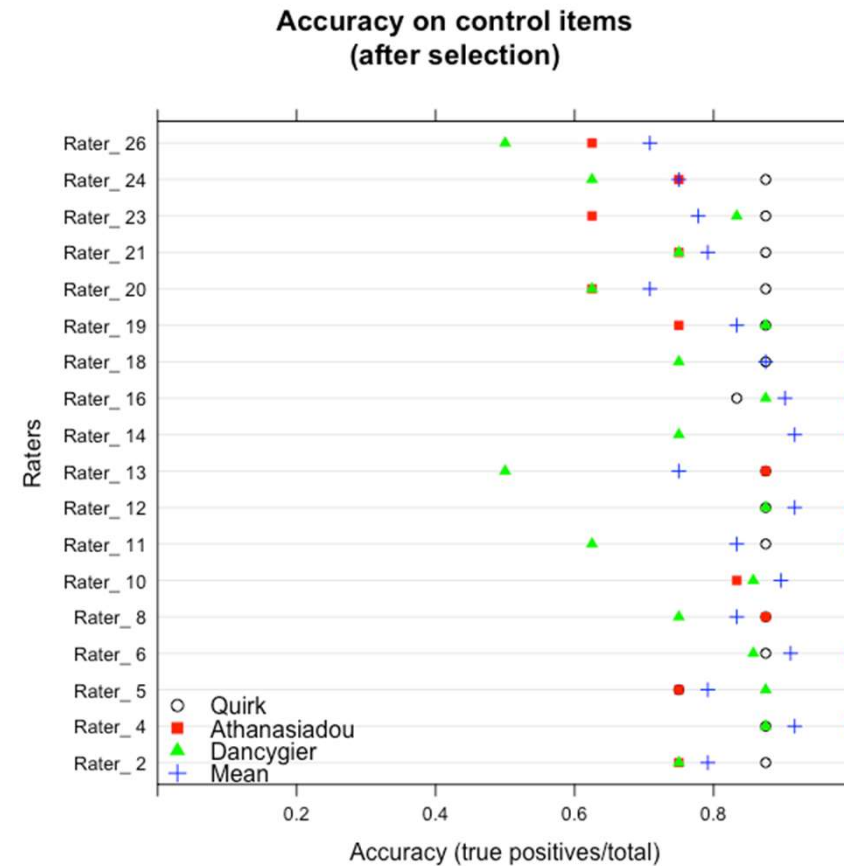
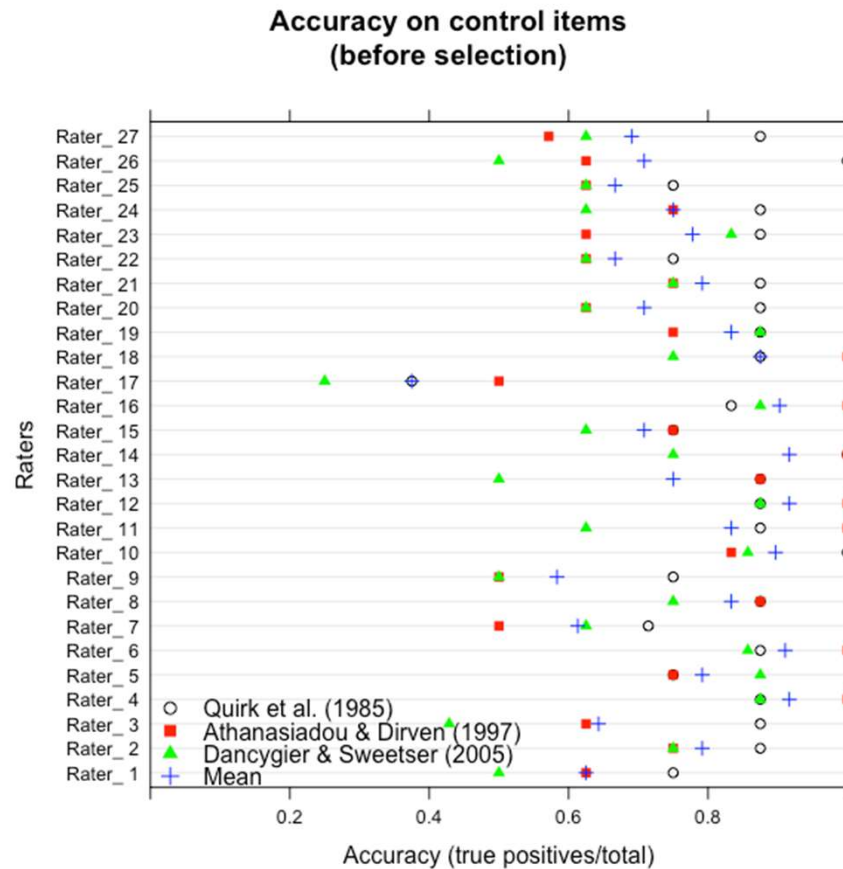
*Mean accuracy per classification (before selection, N=27)*

|                              | Mean accuracy | SD   |
|------------------------------|---------------|------|
| Quirk et al. (1985)          | 0.89          | 0.06 |
| Athanasiadou & Dirven (1997) | 0.83          | 0.15 |
| Dancygier & Sweetser (2005)  | 0.75          | 0.13 |

*Mean accuracy per classification (after selection, N=18)*

# Results: selection of qualified raters

- Distributions of accuracy scores per rater (before and after selection) are presented below.



# Results: inter-rater agreement

|                              | Control | Corpus |
|------------------------------|---------|--------|
| Quirk et al. (1985)          | 0.87    | 0.54   |
| Athanasiadou & Dirven (1997) | 0.59    | 0.33   |
| Dancygier & Sweetser (2005)  | 0.55    | 0.32   |

Inter-rater agreement (Krippendorff's Alpha) per classification for test and corpus items

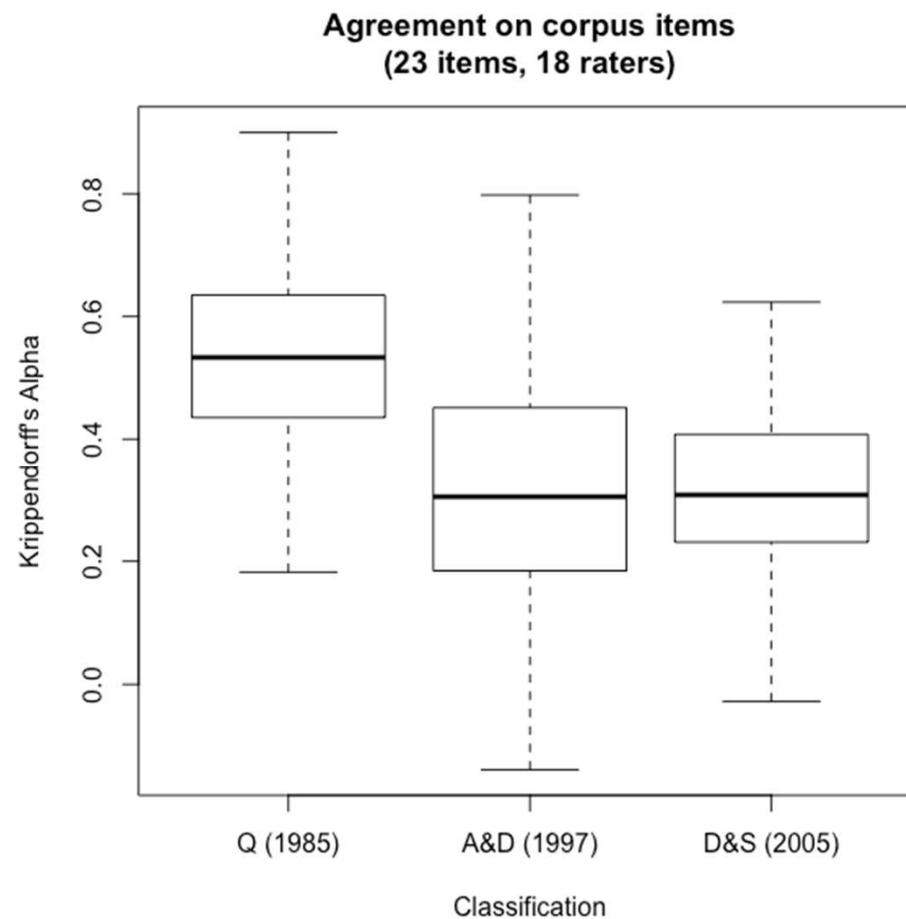
- In comparison to the distribution of agreement coefficients on control items, the raters scored lower on corpus items.
- To allow for a more detailed analysis, a distribution of agreement coefficients was calculated.

# Results: paired-rater agreement

- For all combinations of raters, Krippendorff's Alpha was calculated.
- Low scores, much variation

|                              | Mean | SD   |
|------------------------------|------|------|
| Quirk et al. (1985)          | 0.54 | 0.15 |
| Athanasiadou & Dirven (1997) | 0.33 | 0.18 |
| Dancygier & Sweetser (2005)  | 0.32 | 0.14 |

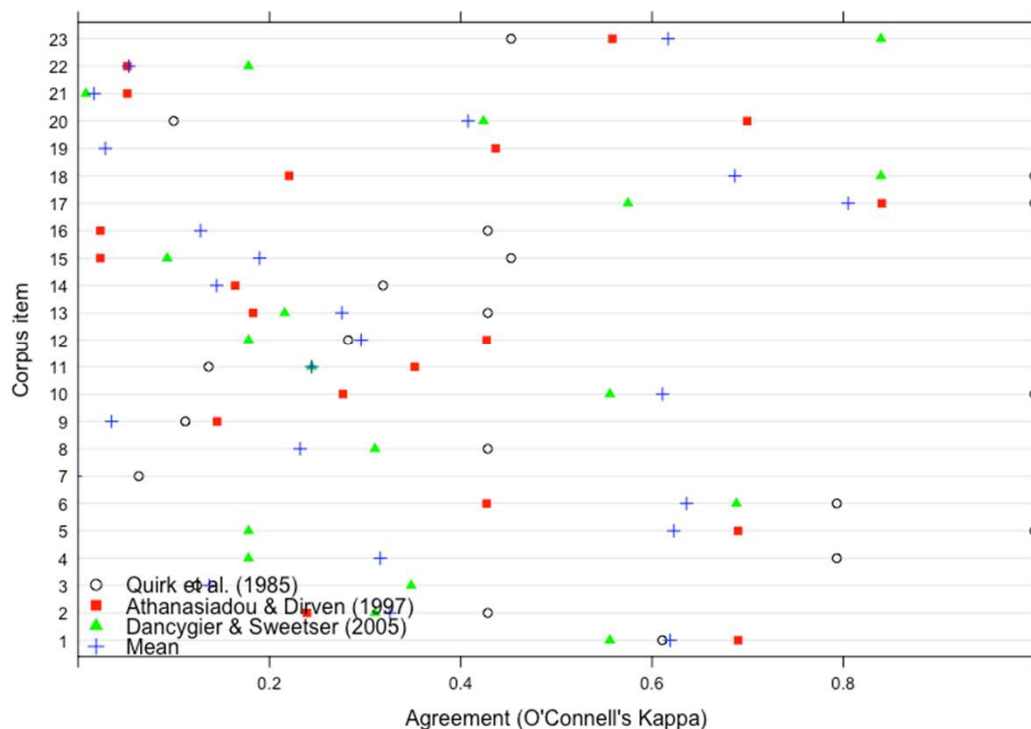
- Landis & Koch (1977)
  - < 0: no agreement
  - 0–0.20: slight agreement
  - 0.21–0.40: fair agreement
  - 0.41–0.60: moderate agreement,
  - 0.61–0.80: substantial
  - 0.81–1: almost perfect agreement



# Results: paired-item agreement

- A combinatorial distribution of ratings per item was generated.
- Average agreement **per item** (O'Connell & Dobson 1984) was calculated.

Agreement on corpus items (18 raters)



23. als je naar Woerleit gaat, aan de overkant van de snelweg is ook een groot bos.
22. ik moet leren delegeren dus als ik iemand de opdracht geef om mijn verslagen te tikken dan delegeer ik toch?
21. als wij nou zaterdag Wietske op gaan halen zaterdagmiddag dan kun jij zaterdagmorgen 't huisje in orde maken.
20. als ik die morgen die floppy's niet kan vinden, waar liggen die dan?
19. als ik dat zo zie dan denk ik ook dat die man 't heeft gedaan.
18. ja maar ik bedoel als dat uh... als je huis een meter in 't water staan dan loopt 't uh toch wel onder.
17. maar als 'k dan 't gas losliet dan sloeg ie af. dus die automatische choke vind ik half.
16. als je 't niet zou weten dan hoor je niet dat de radio aan staat.
15. ja, en als je midden in de fabriek zit ja je weet niet of 't buiten onweert of dondert of regent of de zon schijnt.
14. De verborgen doelstelling is natuurlijk een militaire: want als je een poolstad kan verlichten, kan dat ook met een slagveld.
13. Ja, want de film is goed. En als een film goed is, dan ben jij vanzelf ook goed.
12. 'In volle glorie kan meneer Huisman straks de heronthulling verrichten, als het aan mij ligt', stelt Closset.
11. Het was hem in zijn praktijk opgevallen dat hoeren vaak frigide zijn bij hun klant en dan de smoes verzinnen dat het onprofessioneel zou zijn als ze wel tot een orgasme zouden geraken.
10. Maar als het niet wordt bestraft, gaat zoiets wennen en gaat het steeds een stukje verder.
9. 'OK, als Israël nog onze vijand is voor wat het de Palestijnen aandoet, lees erover, van binnenuit!
8. 'KPN mag ook geen beperkende voorwaarden meer opleggen als het de gegevens verkoopt.
7. Het is gevaarlijk als 'genoeg werk' het enige argument is om aan de Pabo te gaan studeren.
6. Maar als de cd-kopers vaker muziek zouden horen in een goede concertzaal, zou men zich wel eens achter de oren krabben.
5. Nu is het beeld, als u wilt: het vijandbeeld, veel diffuser geworden en de Amerikaanse bereidheid in te grijpen navenant onzekerder.
4. "De vakbonden vergissen zich als ze geloven dat ze sterker zijn dan mijn regering. Ik waarschuw de ZCTU. Ik maak geen grapjes, ik ben bloedserieus."
3. En dan krijg je de drang van: als Jan het kan, moet ik dat ook kunnen.
2. Je kon ze 's nachts om hulp vragen als er stront aan de knikker was...
1. Telkens als ze geestelijk in nood zat, omdat die forehand niet werkte, moest ik haar komen redden.

# Results: problematic cases

- Quirk et al. (1985)
  1. oh . ja ik vind dat wel heel interessant en ik **als ik dat zo zie dan denk ik ook dat die man ‘t heeft gedaan.** ook als je dat hoort maar van de andere kant ja. (19)  
*If I see it like that, I also think that man has done it.*
  2. We moeten oppassen dat de toeloop op de opleidingen in Limburg niet te groot wordt. **Het is gevaarlijk als ‘genoeg werk’ het enige argument is om aan de Pabo te gaan studeren.** (7)  
*It is dangerous if ‘enough work’ is the only argument to study for teacher.*
- Athanasiadou & Dirven (1997)
  1. Met dit oordeel kan een einde komen aan de monopoliepositie van KPN’s CD-foongids, het telefoonboek op cd-rom. **KPN mag ook geen beperkende voorwaarden meer opleggen als het de gegevens verkoopt.** (8)  
*KPN may not impose more restrictive conditions anymore if it sells the data.*
  2. Maar dat kan niet want de ZCTU beschikt niet over de kwaliteiten van een president, aldus Moegabe, die er voor de goede orde aan toevoegde: **“De vakbonden vergissen zich als ze geloven dat ze sterker zijn dan mijn regering.** Ik waarschuw de ZCTU. Ik maak geen grapjes, ik ben bloedserieus.” (4)  
*The unions are mistaken if they believe they are stronger than my government.*

# Results: problematic cases

- Dancygier & Sweetser (2005)
  1. oh . ja ik vind dat wel heel interessant en ik **als ik dat zo zie dan denk ik ook dat die man 't heeft gedaan.** ook als je dat hoort maar van de andere kant ja. (19)  
*If I see it like that, I also think that man has done it.*
  2. mmm ? **als je 't niet zou weten dan hoor je niet dat de radio aan staat.** nee, maar was trouwens wel gaaf dat concert. (16)  
*If you wouldn't know then you would not hear the radio is on.*



# Conclusions: classification of conditionals

- Trained raters were able to classify idealized examples to a high degree for Quirk and Athanasiadou and Dirven's, but less so for Dancygier and Sweetser's classification.
- Trained raters were unable to apply the classification *reliably* to corpus data.
- Low agreement scores may be the result of classification based on interpretation rather than on formal characteristics (cf. Spooren & Degand 2010 on coherence relations).
- It is useful to extend calculating agreement scores to paired-item agreement in order to identify those items that resists classification, here, for instance, conditionals in which the consequent expresses an evaluation of the situation in the antecedent.
- A discussion after the classification task could have enriched the results.
- A bottom-up approach based on more formal characteristics (e.g. clause order, syntactic integration, occurrence of modals, tense backshifting, occurrence of *then*) might be an alternative to more interpretative classification.