The Strong-Weak Pronoun Distinction as a Marker of Literariness

Andreas van Cranenburgh

18 April 2019, Groningen

CLCG Linguistics Lunch

Background: Dutch strong and weak pronouns

	Strong subj, obj	Weak
1st sg	ik, mij	me
2nd sg	jij, jou	je
3rd sg fem	zij, haar	ze
3rd sg masc	hij, hem	-
3rd sg neut	het	-
1st pl	wij, ons	we
2nd pl	jullie	-
3rd pl	zij, hen/hun	ze

Not shown: weak pronouns avoided in written language
or only used as possessive/reflexive.

Cf. Haeseryn et al. (1997). Algemene Nederlandse Spraakkunst

Properties of strong and weak pronouns

Strong: (full, emphatic)

- Often stressed: emphasis/contrast
- Obligatory for:
    - Comparisons (ik ben rijker dan jij, *dan je)
    - Conjunctions of two pronouns (hij en zij)
    - Oblique arguments, eg. rel. clause (voor hen die …)
- Preferred in writing
- Sounds unnatural when repeated

Weak: (reduced, unemphatic)

- Always unstressed: no emphasis
- Obligatory for:
    - Idioms (e.g., dank je, *dank jij)
    - Generic you (je weet maar nooit!)
    - 3rd pers. pl. non-personal (ze, *zij, *hun, *hun);
- Preferred in spoken language
- Can be repeated; often one strong pronoun followed by several weak instances
- Less personal, more informal

Cf. Haeseryn et al. (1997). Algemene Nederlandse Spraakkunst; Bresnan (1998), Markedness and morphosyntactic variation in pronominal

Previous work: why strong vs weak?

Kaiser (2011) and works cited there:

Salience:
null > reduced pronoun > full pronoun > demonstrative > full NP ... etc.

most salient   less salient
referent   referent

Contrast:
The referent is in a contrast relation
to other entities in the discourse

Kaiser (2011). Salience and contrast effects in reference resolution:
The interpretation of Dutch pronouns and demonstratives.

Kaiser (2011)

Kaiser (2011) reports sentence completion
and eye tracking experiments for zij/ze, hij/die

- ► Result: both salience (subject vs object)
  and contrast explain choice of pronoun.

- ► But: data not naturalistic;
  tests salience & contrast effects by manipulating subjects

Kaiser (2011). Salience and contrast effects in reference resolution:
The interpretation of Dutch pronouns and demonstratives.

This talk

- Naturalistic data: novels
- All strong/weak pronouns
- Consider stylistic dimension:
  Weak vs strong pronouns are related to informality and tone

The Riddle of Literary Quality

- ▶ 401 contemporary Dutch-language novels
- ▶ Large reader survey of general public
- ▶ Literary ratings on 7-point Likert scale
- ▶ Stylometry and machine learning with texts of novels

http://literaryquality.huygens.knaw.nl

Research Questions

1. Is there an association between strong/weak pronouns and literariness?
2. To what extent is the choice of pronoun a stylistic matter?

Correlation: pronouns vs literary rating (N=401)
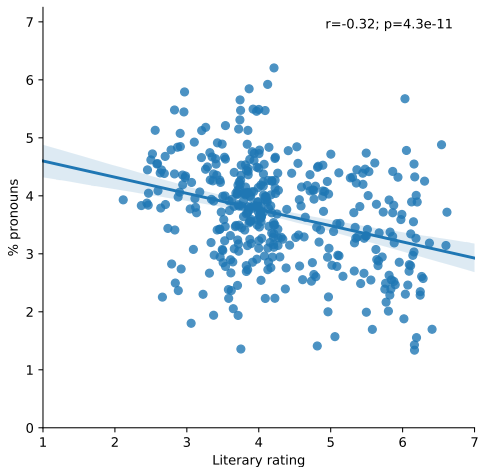
- Count both forms
  Divide by total words
  (rel. freq)
- Less pronouns, more
  literary

Correlation: strong pronouns vs literary rating (N=401)
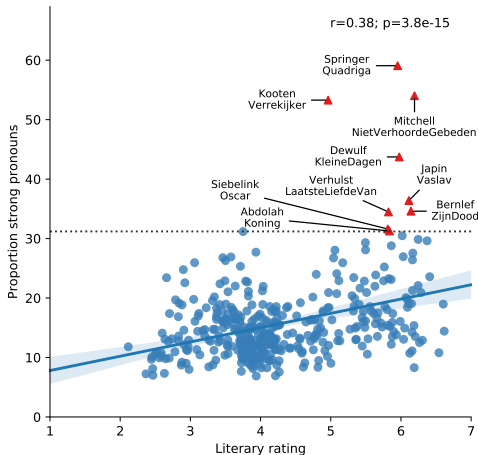
- ▶ Count strong pronouns, divide by count of both forms
- ▶ On average, 85 % of pronouns are weak
- ▶ More strong pronouns, more literary
- ▶ Several strong outliers!

The outliers

Who are these outliers?

- All Dutch authors (except Mitchell)
- All highly literary (> 5)
- Less je, ze-SG,FEM.
- More mij, wij, zij-SG,FEM

Possible explanations

Why are strong pronouns more common in literary texts?

Stylistic choice (deliberate or not):

- Non-literary texts have more informal, idiomatic language
- Literary authors are less afraid of sounding "unnatural"

Discourse structure more complicated:

- Larger number of characters
- Multiple perspective, storylines

give rise to higher frequency of less salient referents and use of contrast.

Manual Analysis

- In first 100 sentences of the outliers
- Annotate each pronoun:
  - Strong vs weak
  - Free choice, preferred, or obligatory
  - Used for emphasis/contrast?
 - Type: personal, possessive, generic, non-personal, verb

Manual Analysis

- In first 100 sentences of the outliers
- Annotate each pronoun:
    - Strong vs weak
    - Free choice, preferred, or obligatory
    - Used for emphasis/contrast?
    - Type: personal, possessive, generic, non-personal, verb

Distribution of types:

Personal	303
Possessive	26
Generic	20
Non-personal	6
Verb	1
Total	356

We'll only consider personal pronouns, which allow both forms.

Breakdown (N=303 pronouns)
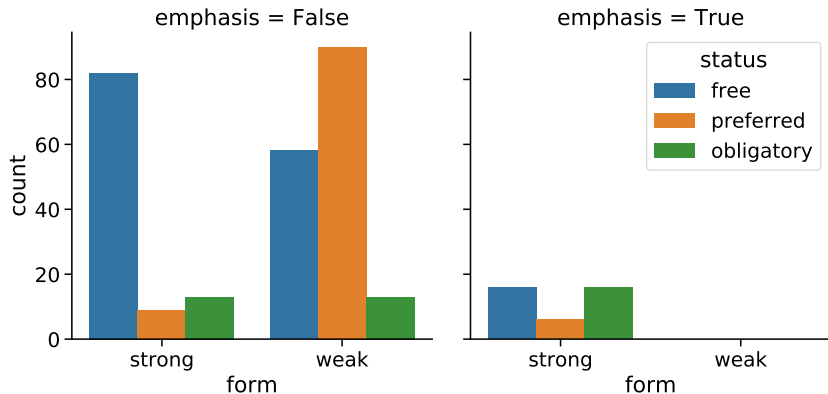
- ▶ Emphasis is rare.
- ▶ Weak often preferred, but large part is free choice.

Typical examples

Weak:

(1) a. Free: We speuren erfgenamen op
   We track down heirs
   b. Preferred: Dat weet je toch?
   You know that right?
   c. Obligatory: Mooie gouvernante is me dat.
   Nice governess that is.

Strong:

(2) a. Free: Hoort u mij?
   Do you hear me?
   b. Preferred: Maar dan kennen ze mij niet.
   But then they haven't met me.
   c. Obligatory: Je ziet dat het niet van mij is!
   You can tell it's not mine!

Interesting examples

Arguably unnatural usage of strong pronoun:

(3) a. Ik keek om mij heen
I looked around me
b. aangezien [. . .] heb ik altijd mijn eigen Duralexglas bij mij
since [. . .] I always have my own Duralex glass with me

Weak vs strong pronouns pick different referents:

(4) Ik heb nooit kunnen vaststellen dat ze mij in de gaten hielden,
al deden ze dat natuurlijk wel, en zij in de eerste plaats.
I have never been able to confirm that they were watching me,
although of course they did , and she most of all.

Conclusion

Answers to Research Questions:

1. Is there an association between strong/weak pronouns and literariness?
   - Negative correlation with # pronouns
   - Positive correlation with proportion strong pronouns
   - Striking outliers: Dutch literary authors love strong pronouns

Conclusion

Answers to Research Questions:

1. Is there an association between strong/weak pronouns and literariness?
   - Negative correlation with # pronouns
   - Positive correlation with proportion strong pronouns
   - Striking outliers: Dutch literary authors love strong pronouns
2. To what extent is the choice of pronoun a stylistic matter?
   - Majority is free, stylistic choice.
   - Obligatoriness and emphasis are rare.