

The Strong-Weak Pronoun Distinction as a Marker of Literariness

Andreas van Cranenburgh

28 February 2019, Würzburg

Digital Stylistics in Romance Studies and Beyond conference

Outline

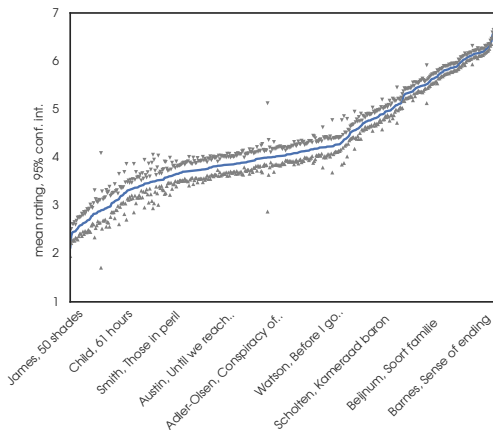
- ▶ General problem: What is literature?
- ▶ This talk: Strong/weak pronouns
- ▶ Results
 - ▶ Correlations
 - ▶ Outliers
 - ▶ Analysis

Background: What is literature?

What makes a literary novel *literary*?

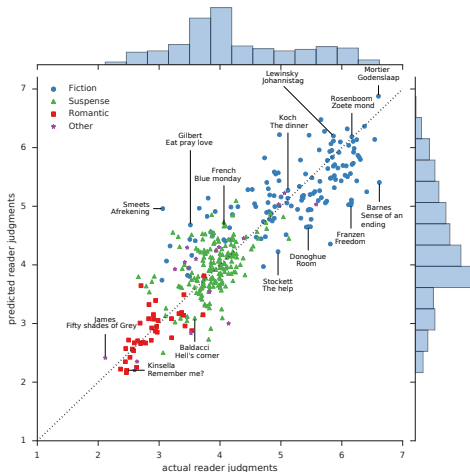
- ▶ Cultural capital of critics and publishers
- ▶ Subjective aesthetic value-judgments
- ▶ “Objective” textual features, writing style: *literariness*

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

Previous results: 76.0 % R^2



- ▶ Literariness is highly predictable from text
- ▶ Word frequencies, clichés, syntactic complexity yield good predictions,
- ▶ but many many features, hard to interpret.

This talk is not about getting better predictions, but about understanding one specific stylistic aspect

van Cranenburgh & Bod (2017). A data-oriented model of literary language
van Cranenburgh et al (2019). Vector space explorations of literary language

Background: strong and weak pronouns

- ▶ Some Dutch pronouns have strong and weak (reduced) forms
- ▶ Same meaning, but strong/weak can be obligatory, preferred, or free choice

	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.

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

Missing: stylistic dimension

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

Research Questions

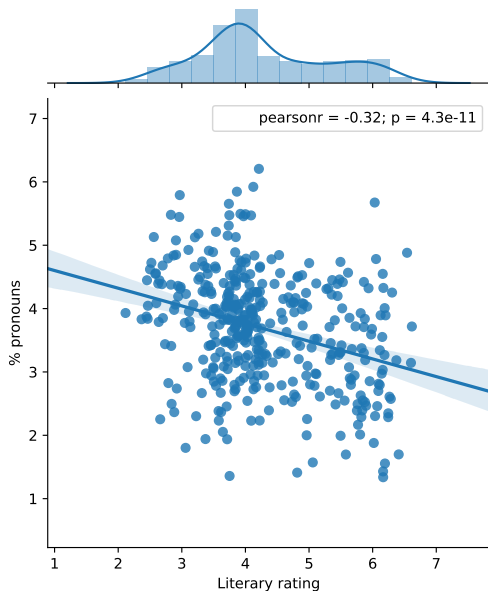
1. Is there an association between strong/weak pronouns and literariness?
2. How common is it for these variants to be
 - a Free stylistic choice
 - b Preferred
 - c Obligatory
3. Why could there be such an association?

Method

Calculate correlation between two variables:

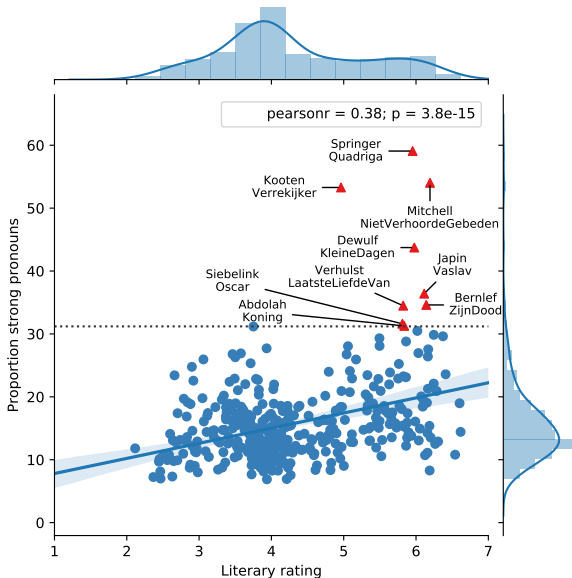
1. Mean literary ratings of novels
2. Two independent measurements:
 - a Baseline: frequency of both pronoun forms
 - b Main result: proportion of strong pronouns vs both forms

Correlation: pronouns vs literary rating (N=401)



- ▶ Count both forms
- ▶ Divide by total words (rel. freq)
- ▶ Result:
Less pronouns, more literary
- ▶ Probably proxy for
amount of dialogue vs
narrative description

Correlation: strong pronouns vs literary rating (N=401)



- ▶ Count strong pronouns, divide by count of both forms
- ▶ Independent of total number of pronouns
- ▶ On average, 85 % of pronouns are weak
- ▶ More strong pronouns, more literary
- ▶ Several strong outliers!

Distribution of pronouns in outliers

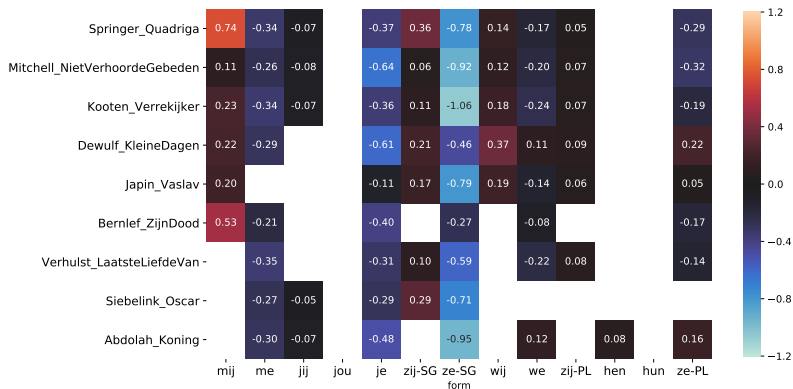


Table: Divergence of relative frequencies (wrt corpus mean)

- ▶ All Dutch authors (except Mitchell), 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

- ▶ Look at pronouns in first 100 sentences of the outliers
- ▶ Annotate:
 - ▶ Strong vs weak
 - ▶ Free choice, preferred, or obligatory
 - ▶ If strong, used for emphasis?
 - ▶ Type: personal, possessive, generic, non-personal, verb

Manual Analysis

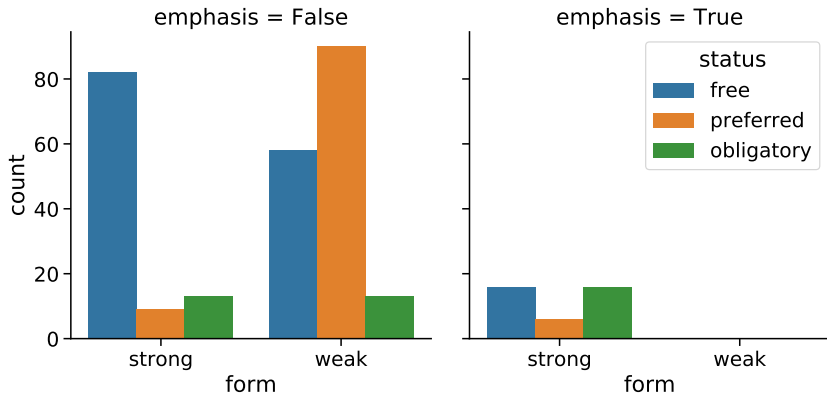
- ▶ Look at pronouns in first 100 sentences of the outliers
- ▶ Annotate:
 - ▶ Strong vs weak
 - ▶ Free choice, preferred, or obligatory
 - ▶ If strong, used for emphasis?
 - ▶ Type: personal, possessive, generic, non-personal, verb

Distribution of types:

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

For the rest, 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 between pronouns and literariness
 - ▶ Positive correlation between strong pronouns and literariness
 - ▶ Striking outliers: Dutch authors who use lots of strong pronouns

Conclusion

Answers to Research Questions:

1. Is there an association between strong/weak pronouns and literariness?
 - ▶ Negative correlation between pronouns and literariness
 - ▶ Positive correlation between strong pronouns and literariness
 - ▶ Striking outliers: Dutch authors who use lots of strong pronouns
2. How common is it for these variants to be free choice, preferred, or obligatory?
 - ▶ Majority is free, stylistic choice.
 - ▶ Even in the outliers, weak pronouns are often preferred
 - ▶ Obligatoriness and emphasis are rare.

Conclusion

Answers to Research Questions:

1. Is there an association between strong/weak pronouns and literariness?
 - ▶ Negative correlation between pronouns and literariness
 - ▶ Positive correlation between strong pronouns and literariness
 - ▶ Striking outliers: Dutch authors who use lots of strong pronouns
2. How common is it for these variants to be free choice, preferred, or obligatory?
 - ▶ Majority is free, stylistic choice.
 - ▶ Even in the outliers, weak pronouns are often preferred
 - ▶ Obligatoriness and emphasis are rare.
3. Why could there be such an association?
 - ▶ Seems to be predominantly a stylistic choice
 - ▶ Should look into discourse context

EXTRA SLIDES

Future Work

- ▶ Analyze antecedents: distance, grammatical function etc.
- ▶ See if weak/strong distinction is useful as a feature for automatic anaphora resolution
- ▶ Compare with non-literary corpora
- ▶ etc.

Properties of strong and weak pronouns

Strong: (emphatic)

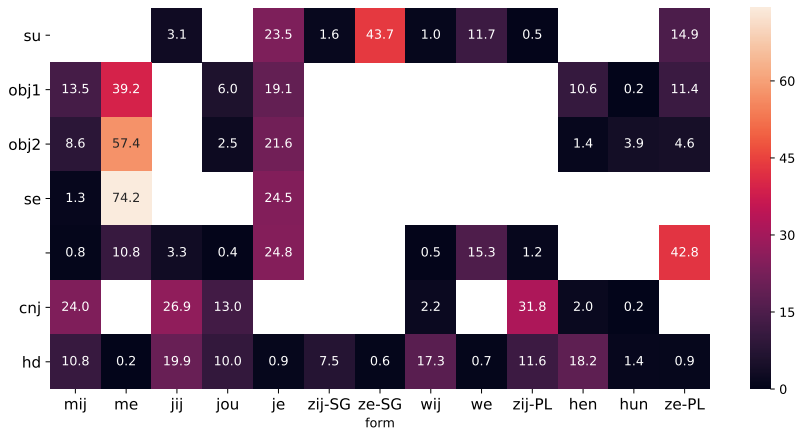
- ▶ Default, unmarked
- ▶ Often stressed: emphasis/contrast
- ▶ Less salient referents
- ▶ Obligatory for:
 - ▶ Comparisons
(ik ben rijker dan **jij**, *dan **je**)
 - ▶ Conjunctions of two pronouns
(**hij** en **zij**)
 - ▶ Oblique arguments, eg. rel. clause (voor **hen** die ...)
- ▶ Preferred:
 - ▶ Sounds unnatural when repeated
 - ▶ Preferred in writing, even when weak form used when same sentence is spoken (applies to **ie**, 'm, d'r, 't, but also to a lesser extent for **me/je/we/ze**)

Weak: (unemphatic)

- ▶ Reduced, marked
- ▶ Always unstressed: no emphasis
- ▶ Salient referents
- ▶ 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:
 - ▶ Can be repeated; or sentence has 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 systems

Distribution of grammatical functions



Percentages add up to 100 for each row