UNIVERSITEIT GENT

FACULTY OF ARTS AND PHILOSOPHY

Koen Plevoets

# INFORMATIONAL LOAD AS A TRIGGER FOR DISFLUENCIES IN INTERPRETING

# Overview

Intro

Research question

Data

Method

Analysis

Conclusion


EUROPEAN PARLIAMENT

# Intro

Interpreting: The rendition of utterances in another language

- Consecutive
- Simultaneous
- …

‘Online’ translation

# Intro

‘Interpreting is a cognitively demanding activity’

# Intro

‘Interpreting is a cognitively demanding activity’

- Multitasking: Division of attention to different concurring tasks
- ‘Tightrope hypothesis’: Interpreters work at the limits of their processing capacities

(Gile 1999)

# Intro

## Effort Model

(Gile 1985; 1997)

Interpreting= L + P + M + C

- L: Listening effort
- P: Production effort
- M: Memory effort
- C: Coordination effort

Gerver (1976)

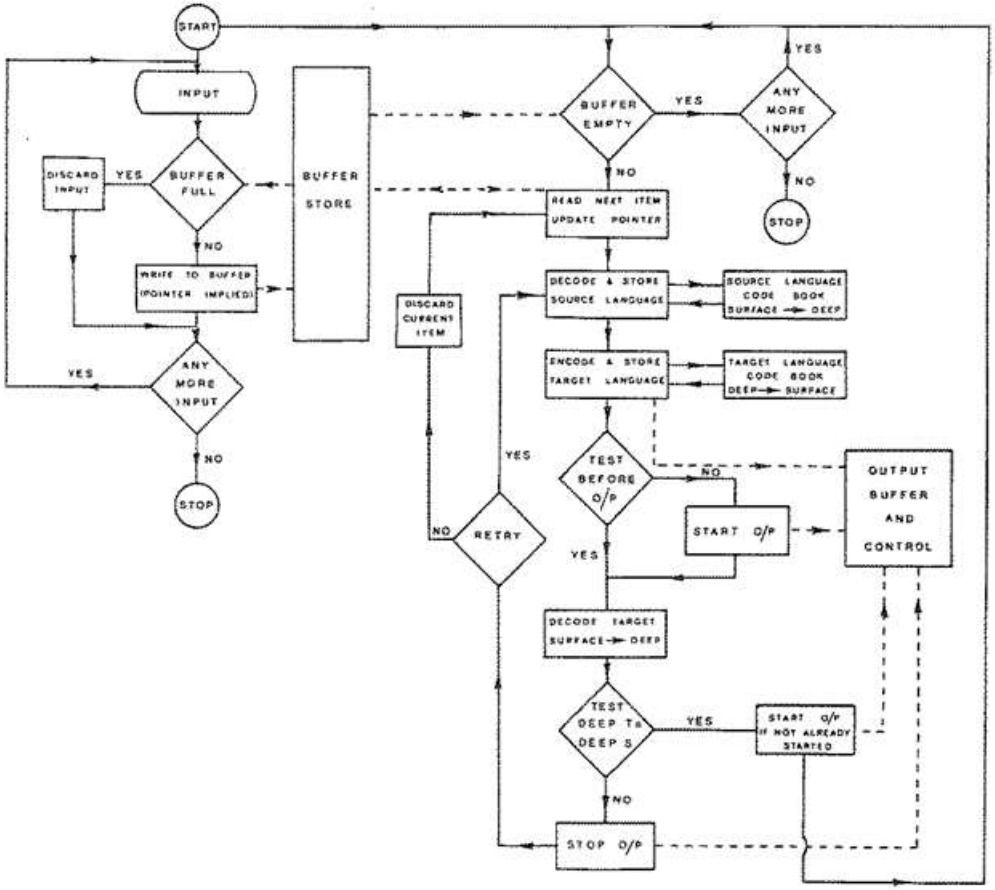
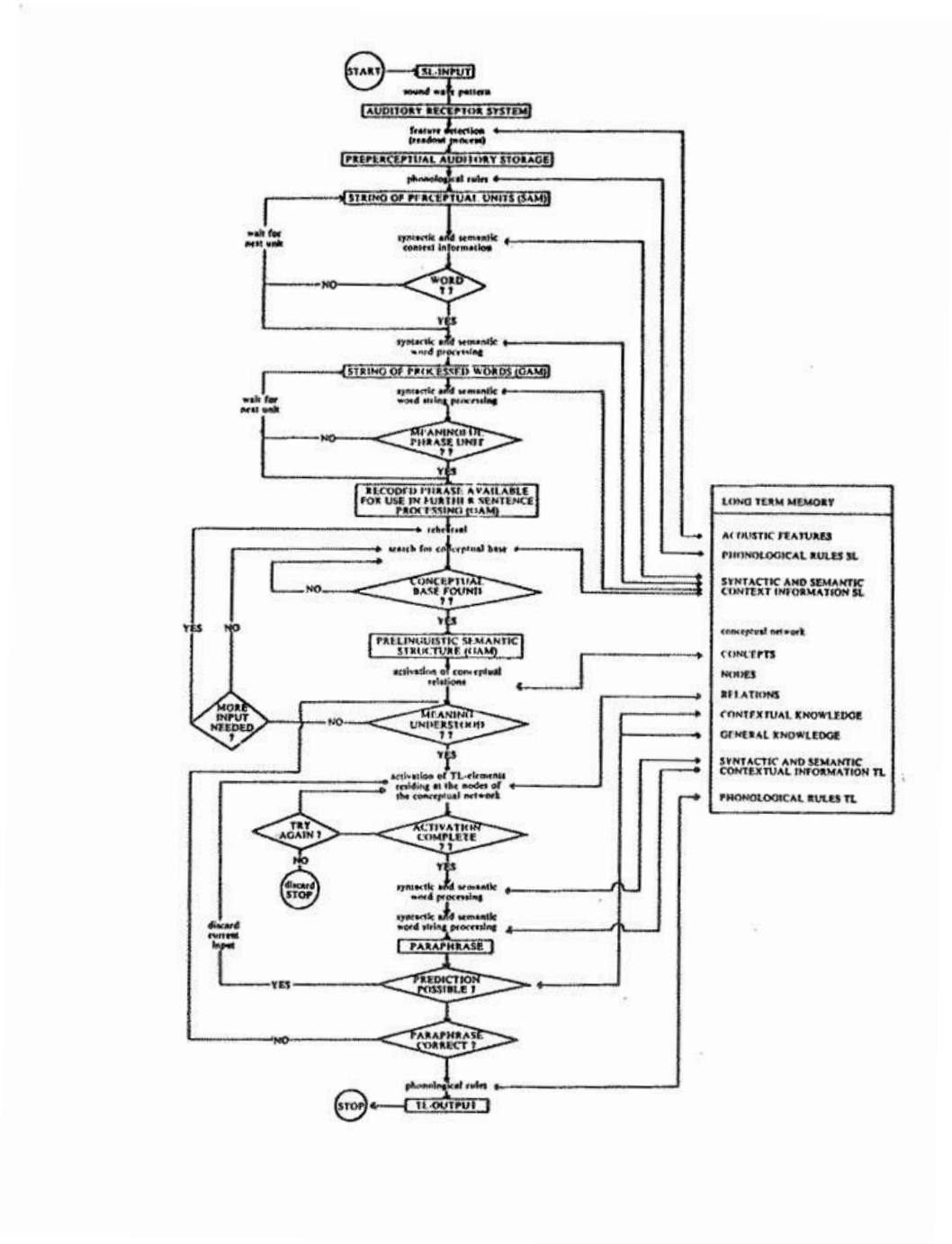
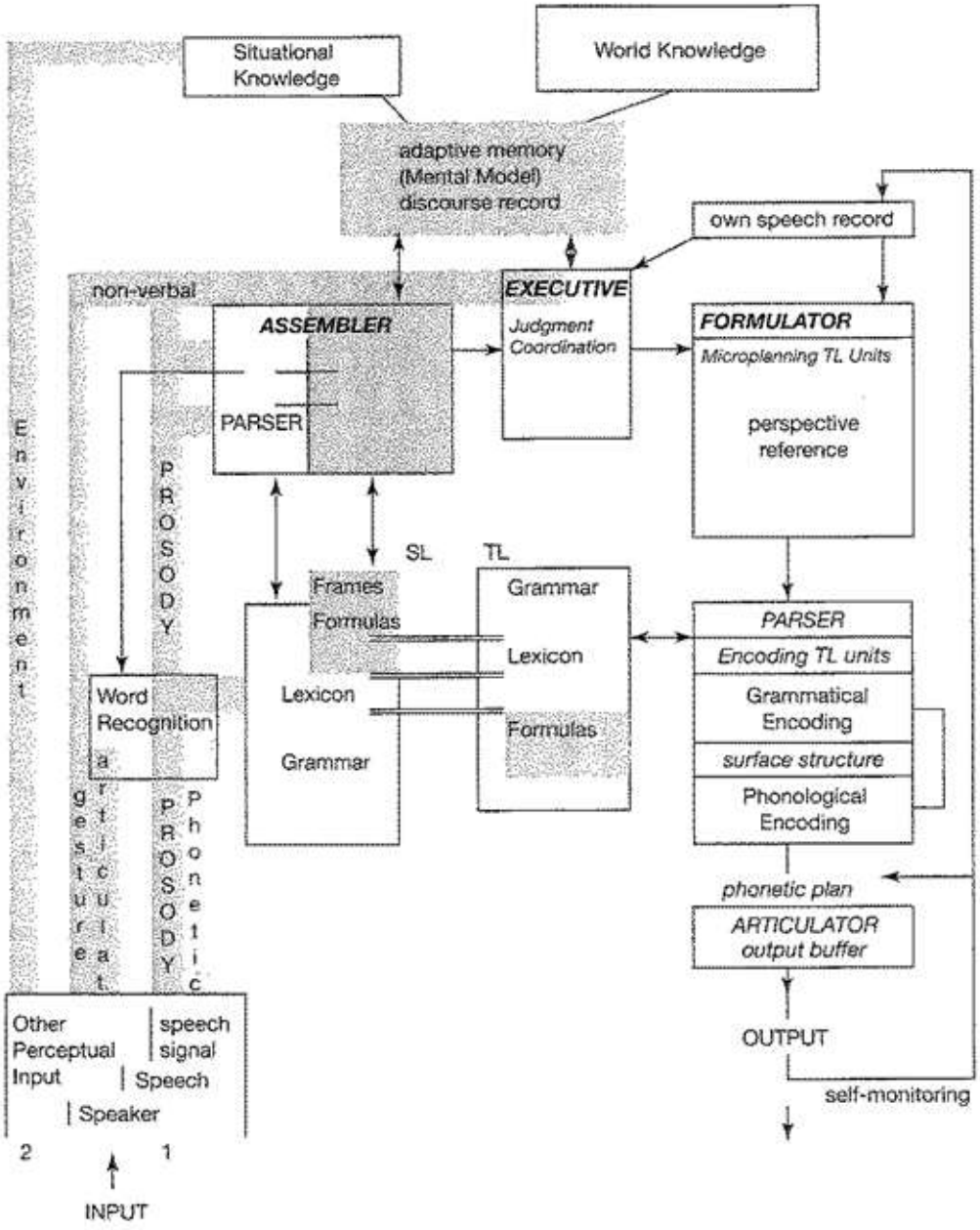



Figure 1: A model of the simultaneous interpretation process.

Moser (1978)


START
SL-INPUT
sound wave pattern
AUDITORY RECEPTOR SYSTEM
feature detection (readout process)
PREPERCEPTUAL AUDITORY STORAGE
phonological rules
STRING OF PERCEPTUAL UNITS (SAM)
wait for next unit
syntactic and semantic context information
NO
WORD ??
YES
syntactic and semantic word processing
STRING OF PROCESSED WORDS (SAM)
wait for next unit
syntactic and semantic word string processing
NO
MEANINGFUL PHRASE UNIT ??
YES
RECODED PHRASE AVAILABLE FOR USE IN FURTHER SENTENCE PROCESSING (SAM)
rehearsal
search for conceptual base
NO
CONCEPTUAL BASE FOUND ??
YES
YES
NO
PRELINGUISTIC SEMANTIC STRUCTURE (SAM)
activation of conceptual relations
MORE INPUT NEEDED ?
NO
MEANING UNDERSTOOD ??
YES
activation of TL-elements residing at the nodes of the conceptual network
TRY AGAIN ?
ACTIVATION COMPLETE ??
NO
YES
discard STOP
syntactic and semantic word processing
discard current input
syntactic and semantic word string processing
PARAPHRASE
YES
PREDICTION POSSIBLE ?
NO
PARAPHRASE CORRECT ?
phonological rules
STOP
TL-OUTPUT
LONG TERM MEMORY
ACOUSTIC FEATURES
PHONOLOGICAL RULES SL
SYNTACTIC AND SEMANTIC CONTEXT INFORMATION SL
conceptual network
CONCEPTS
NODES
RELATIONS
CONTEXTUAL KNOWLEDGE
GENERAL KNOWLEDGE
SYNTACTIC AND SEMANTIC CONTEXTUAL INFORMATION TL
PHONOLOGICAL RULES TL

Setton (1999)


Situational Knowledge
World Knowledge
adaptive memory (Mental Model) discourse record
own speech record
non-verbal
EXECUTIVE
Judgment Coordination
ASSEMBLER
PARSER
FORMULATOR
Microplanning TL Units
perspective reference
Environment
PROSODY
SL
TL
Frames
Formulas
Lexicon
Grammar
Grammar
Lexicon
Formulas
Word Recognition
articulatory gesture
PROSODY
Phonetic
PARSER
Encoding TL units
Grammatical Encoding
surface structure
Phonological Encoding
phonetic plan
ARTICULATOR output buffer
OUTPUT
self-monitoring
Other Perceptual Input
speech signal
Speech
Speaker
2
1
INPUT

# Intro

Research into cognitive load in interpreting:

| | **Temporal characteristics** | **Textual characteristics** |
|---|---|---|
| Source text | Speech rate | Complexity |
| Target text | Ear-Voice Span | Disfluencies |

# Intro

## Temporal characteristics:

- Speech rate: 120 words/minute is comfortable
  (Gerver 1969)
- Ear-Voice Span/“décalage”: 2-3 seconds on average
  (Treisman 1965; Anderson 1994)

# Intro

## Textual characteristics:

- Complexity: Cognitive load increases with
  - Lexical content (Gile 1995)
  - Numbers (Gile 1995)
  - Syntactic embedding (Dillinger 1994; Tommola & Helevä 1998)
- Disfluencies: e.g. silent/filled pauses: *uh(m)*,…

  (Tissi 2000; Mead 2002; Bakti 2009; Tóth 2011)

# Intro

Setton (1999: 247)

| | **Attention to input** | **Attention to formulation** |
|---|---|---|
| Long silent pause | High | - |
| Short pausing | Normal listening | Routine planning |
| Filled pause | Normal listening | Routine planning |
| Mixed: Short & filled pauses & voice effects | Normal listening | Routine planning |
| Long filled pause | Relaxed or off | Planning/Searching |
| Fluent unmodulated string | Relaxed or off | Off |

# Intro

‘Disfluencies are a window on cognitive planning’

(Arnold et al. 2003; Bortfeld et al. 2001; Clark & Fox Tree 2002; Corley & Stewart 2008; Watanabe et al. 2008)

# Research question

To what extent do disfluencies in interpreting depend on informational complexity?

- Lexical content
- Numbers
- Syntactic embedding

# Research question

To what extent do disfluencies in interpreting depend on informational complexity?

- Lexical content
- Numbers
- Syntactic embedding

# Data

European Parliament Interpreting Corpus – Ghent

Plenary sessions of the European Parliament

2006-2008

French, Spanish, Dutch, and English

190 000 tokens… and rising

# Data

European Parliament Interpreting Corpus – Ghent

Transcribed according to VALIBEL-corpus (Bachy et al. 2007)

POS-tagged and chunked by means of LeTs (Van de Kauter et al. 2013)

# Data

Reference corpus: Spoken Dutch Corpus
(Oostdijk 2000)

Component g: Parliamentary debates

POS-tagged

10 million tokens

- Flanders: 1/3
- The Netherlands: 2/3

# Data

Reference corpus: Spoken Dutch Corpus
(Oostdijk 2000)

Component g: Parliamentary debates

360 000 tokens

- Flanders: 140 000
- The Netherlands: 220 000

# Data

| | Nr. of files | Nr. of sentences |
|---|---|---|
| EPICg – FRA (source) | 108 | 1458 |
| EPICg – DUT (target) | 108 | 1437 |
| SDCfl | 155 | 8293 |
| SPCnl | 85 | 10753 |

# Method

Per sentence:

- Nr. of *uh(m)*
- Nr. of content words
- Nr. of numerals
- Nr. of function words (= remainder)
- …

# Method

Predict Nr. of *uh(m)* on the basis of content words, numerals, function words AND 'language'

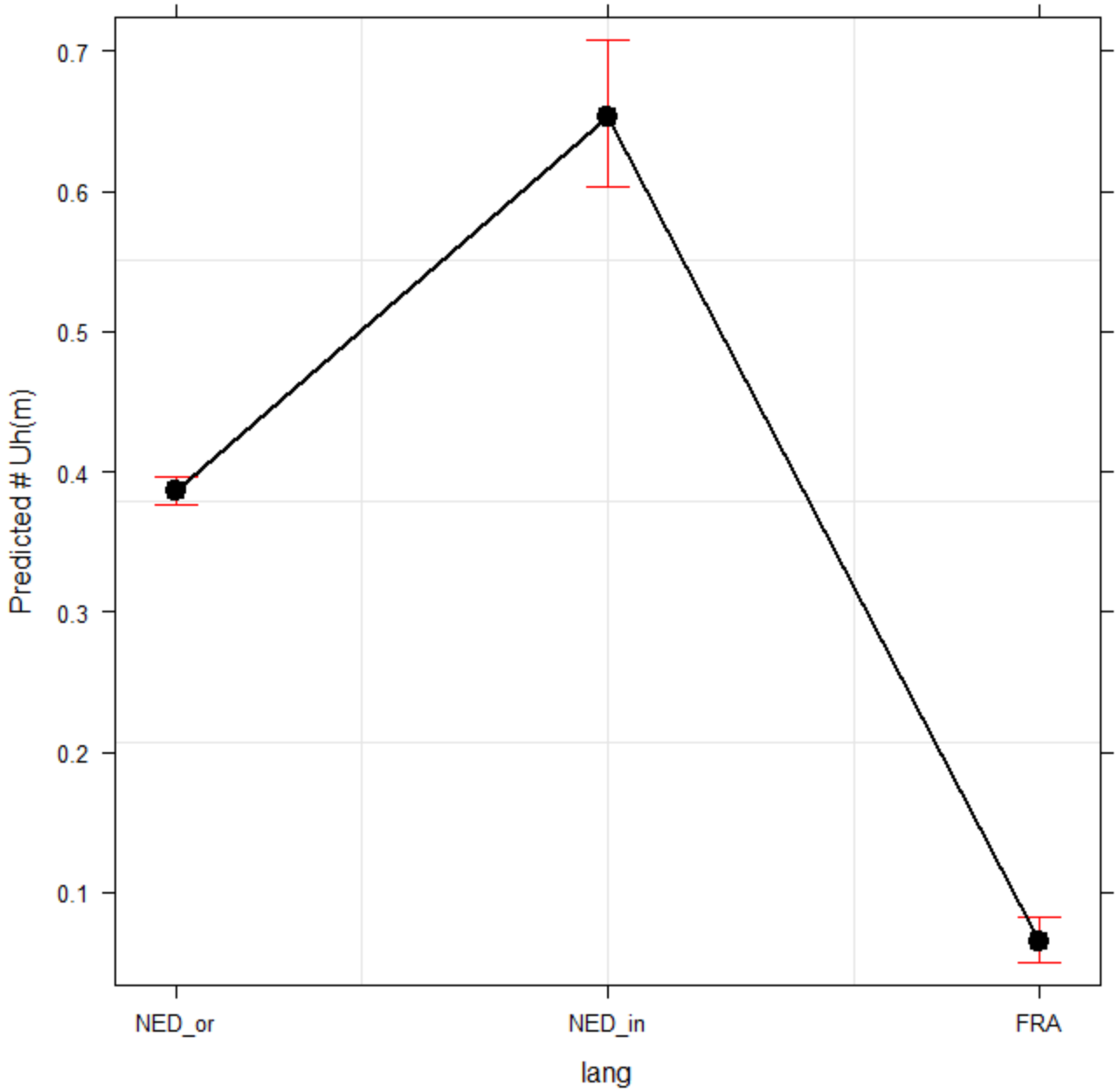
i.e. non-interpreted Dutch, interpreted Dutch, and French source

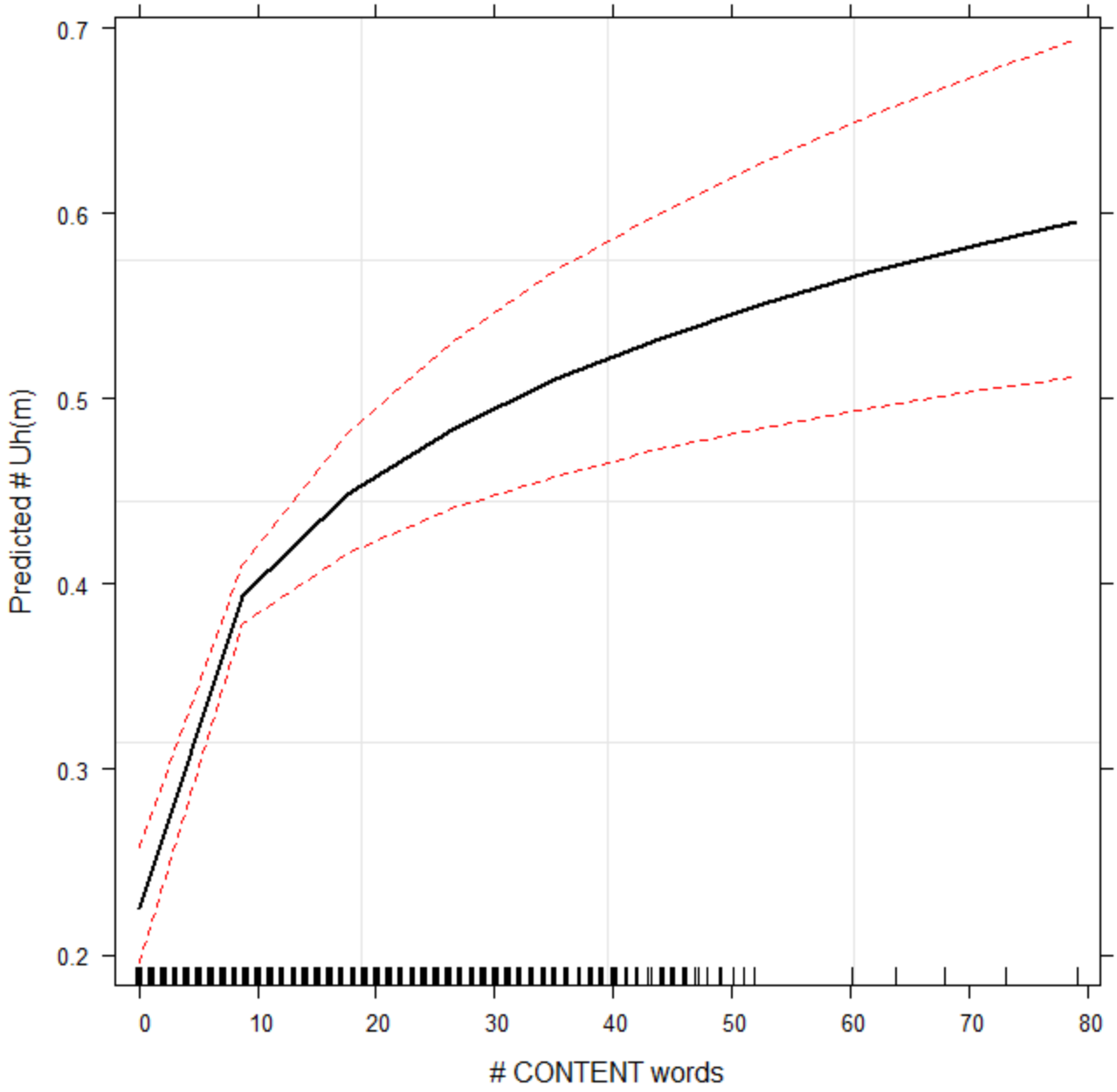
Poisson regression

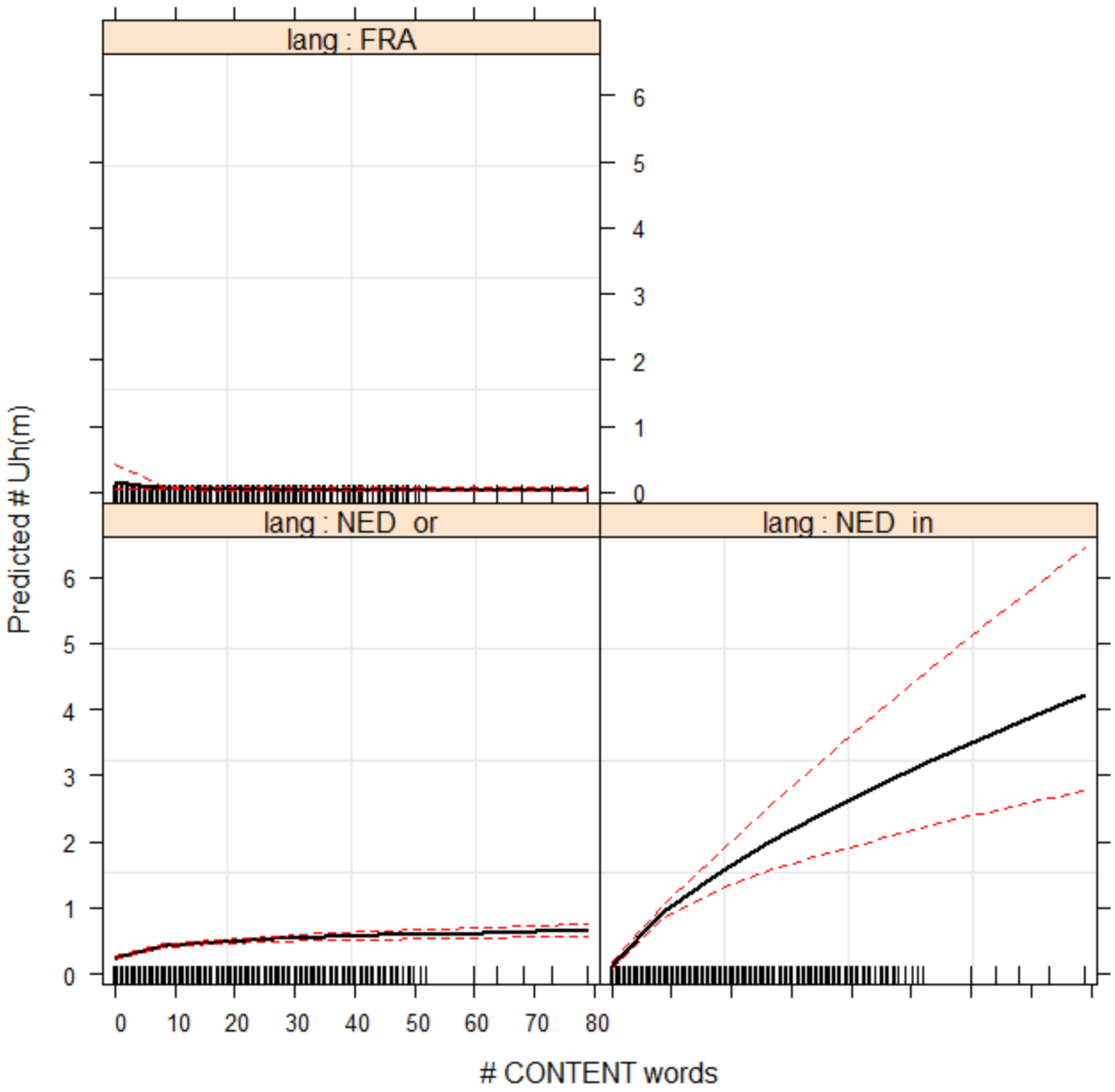
(Verified with Robust regression)

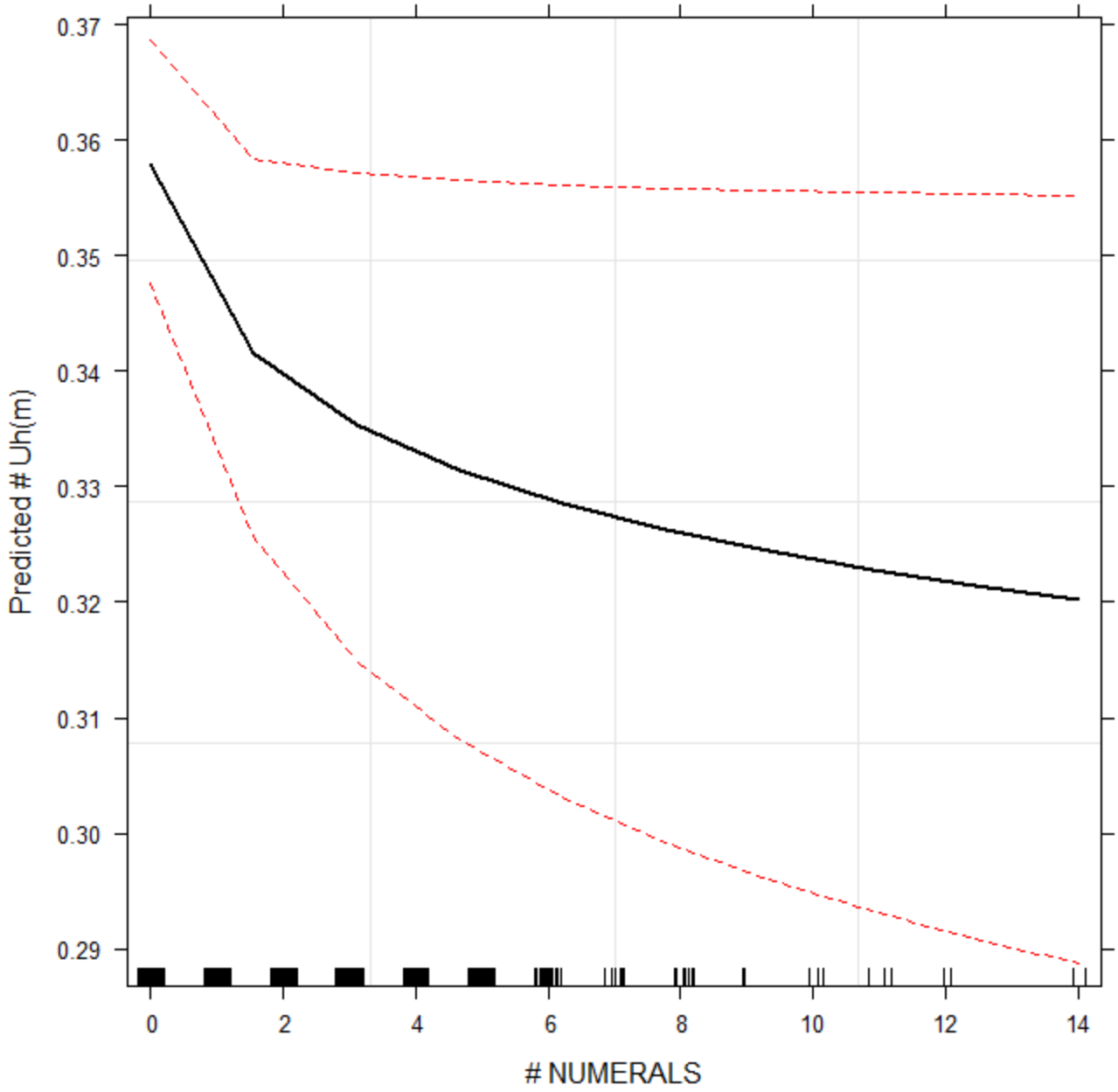
# Analysis

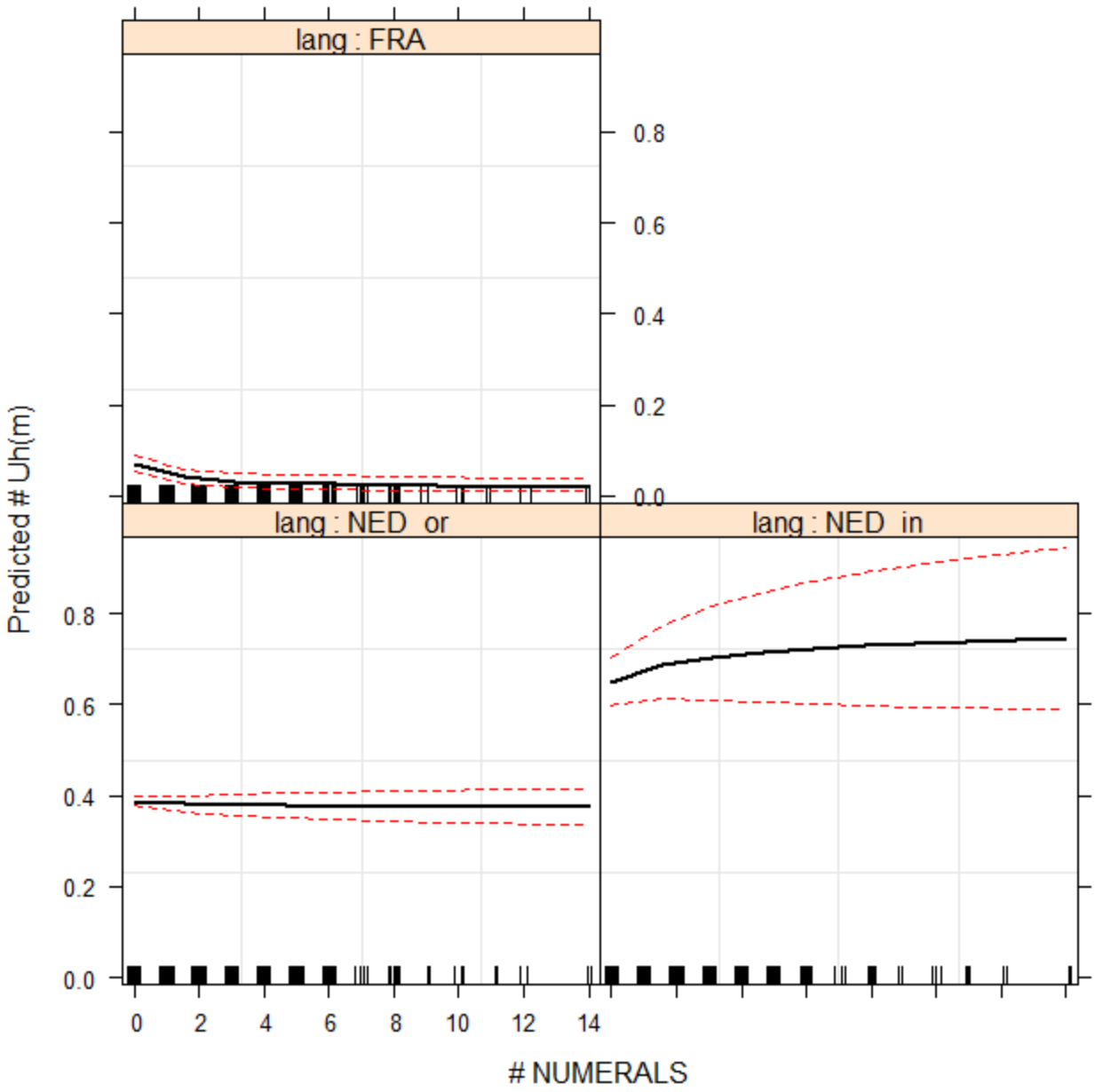
1. At the level of the sentences
2. To measure the effect of the French source load on the Dutch interpretations: At the level of the files

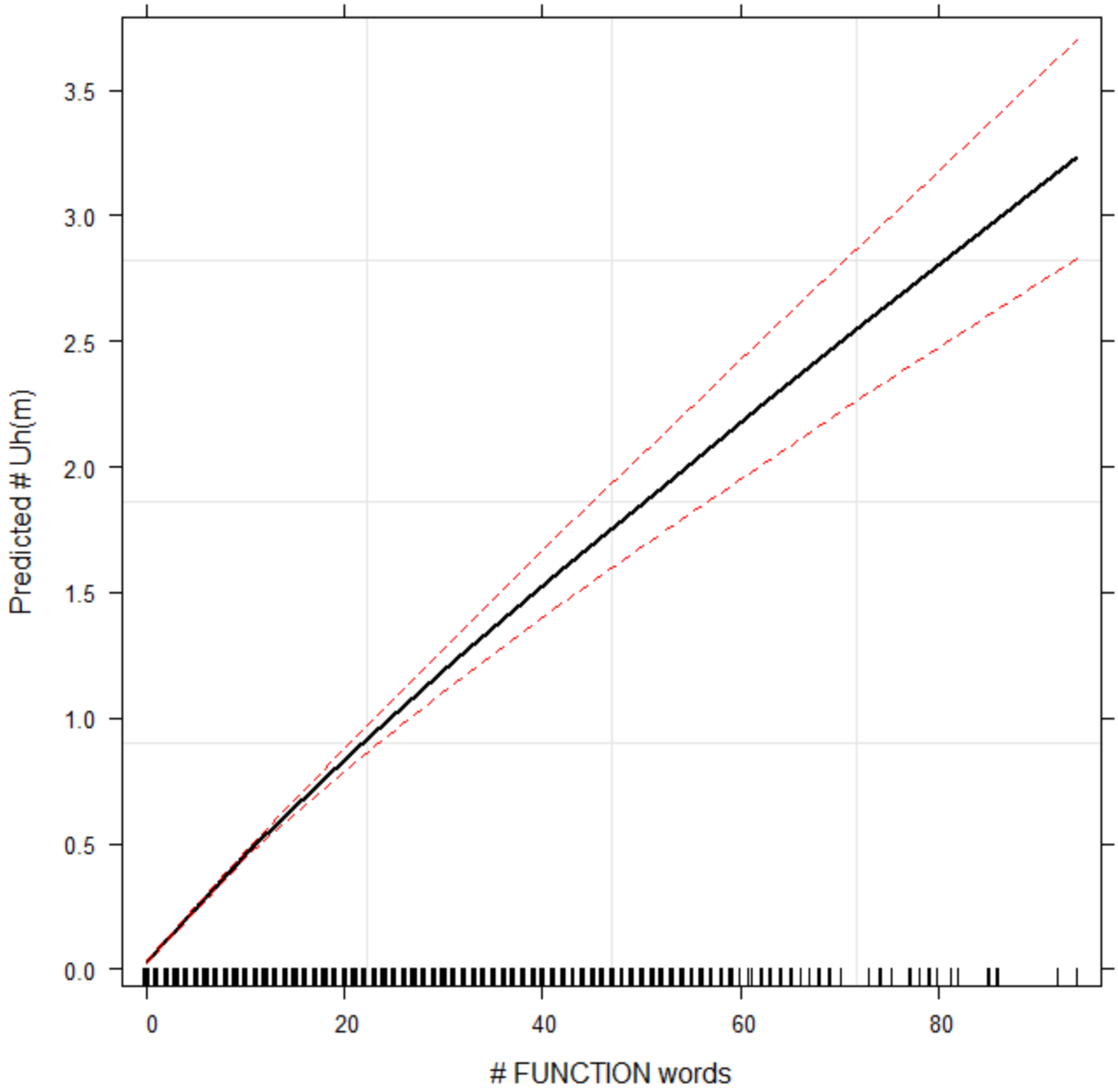

0.7
0.6
0.5
0.4
0.3
0.2
0.1
Predicted # Uh(m)
NED_or
NED_in
FRA
lang

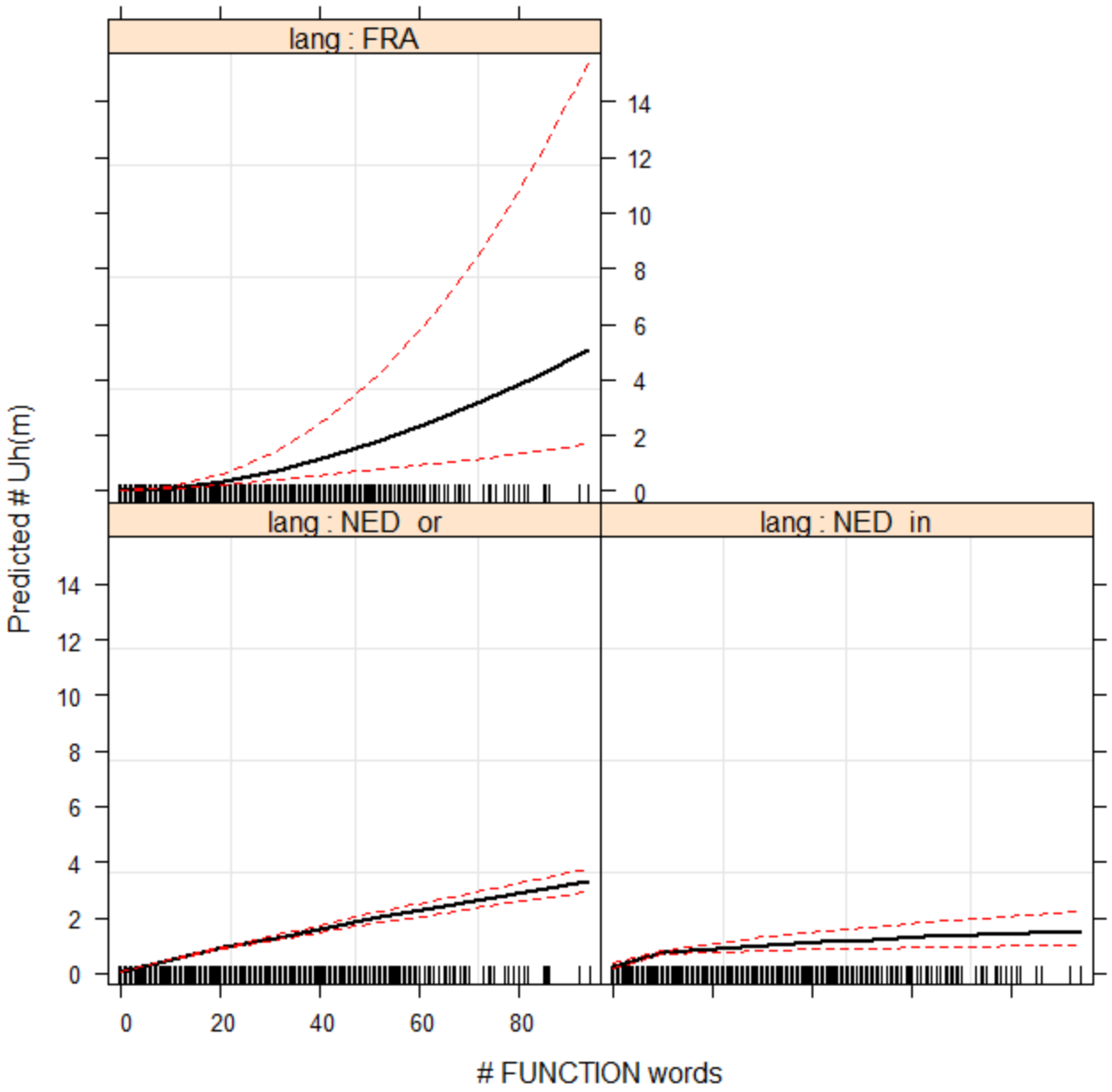

0.7
0.6
0.5
0.4
0.3
0.2
Predicted # Uh(m)
0
10
20
30
40
50
60
70
80
# CONTENT words


lang : FRA
lang : NED or
lang : NED in
Predicted # Uh(m)
# CONTENT words
0
1
2
3
4
5
6
0
10
20
30
40
50
60
70
80


0.37
0.36
0.35
0.34
0.33
0.32
0.31
0.30
0.29
Predicted # Uh(m)
0
2
4
6
8
10
12
14
# NUMERALS

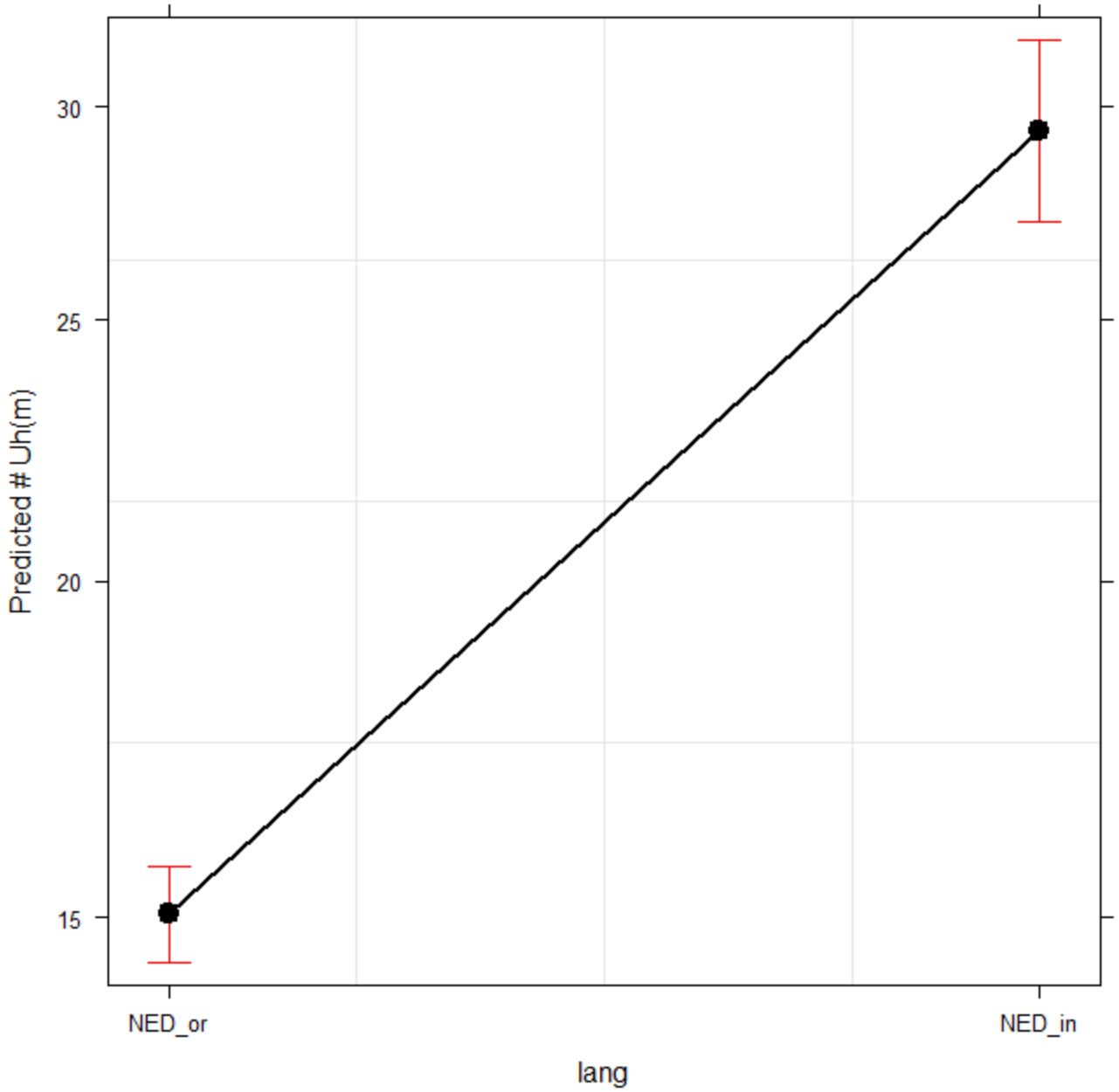

lang : FRA
lang : NED or
lang : NED in
Predicted # Uh(m)
0.0
0.2
0.4
0.6
0.8
0
2
4
6
8
10
12
14
# NUMERALS

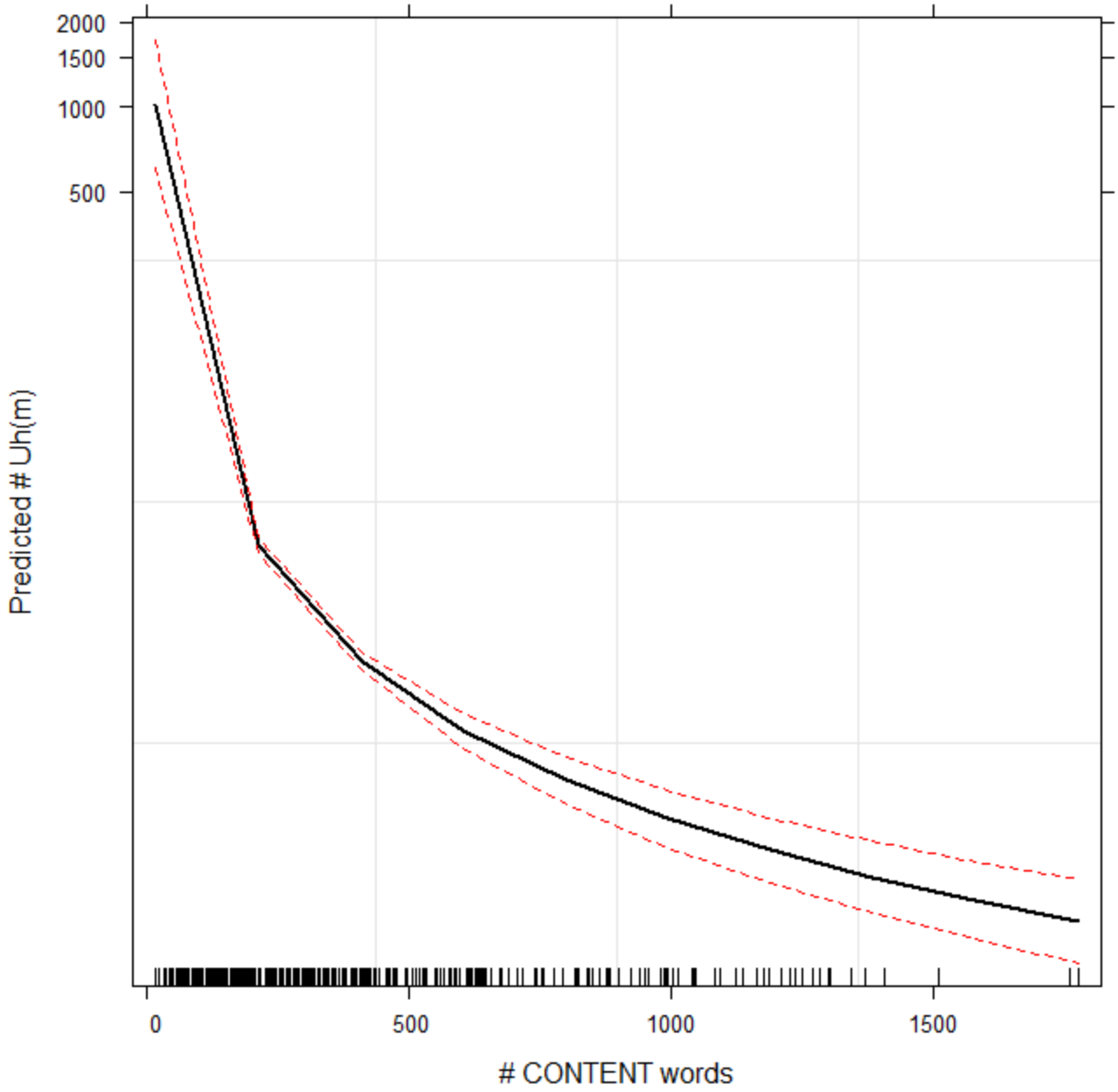

Predicted # Uh(m)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
20
40
60
80
# FUNCTION words

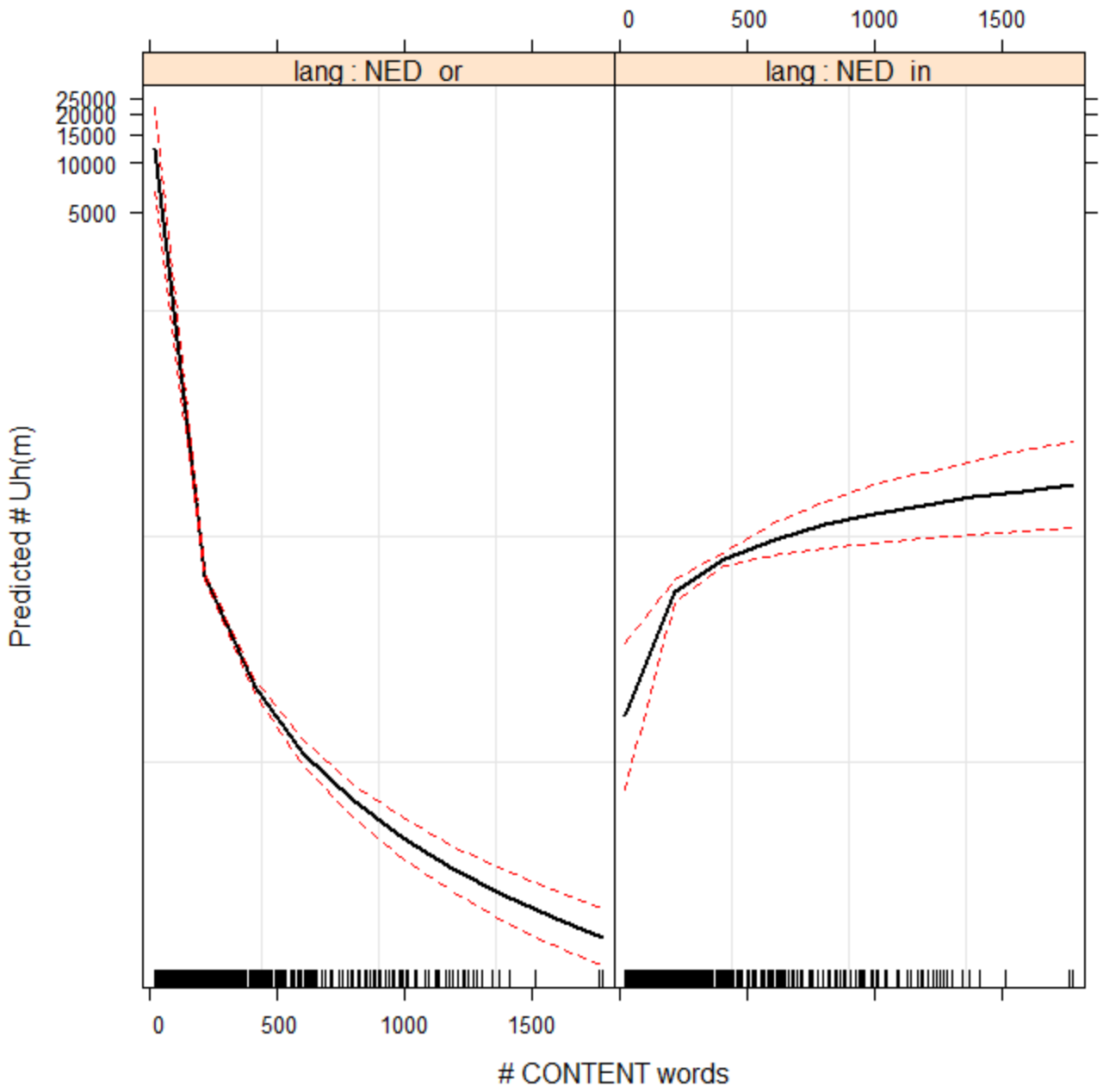

lang : FRA
lang : NED or
lang : NED in
Predicted # Uh(m)
# FUNCTION words
0
2
4
6
8
10
12
14
0
20
40
60
80

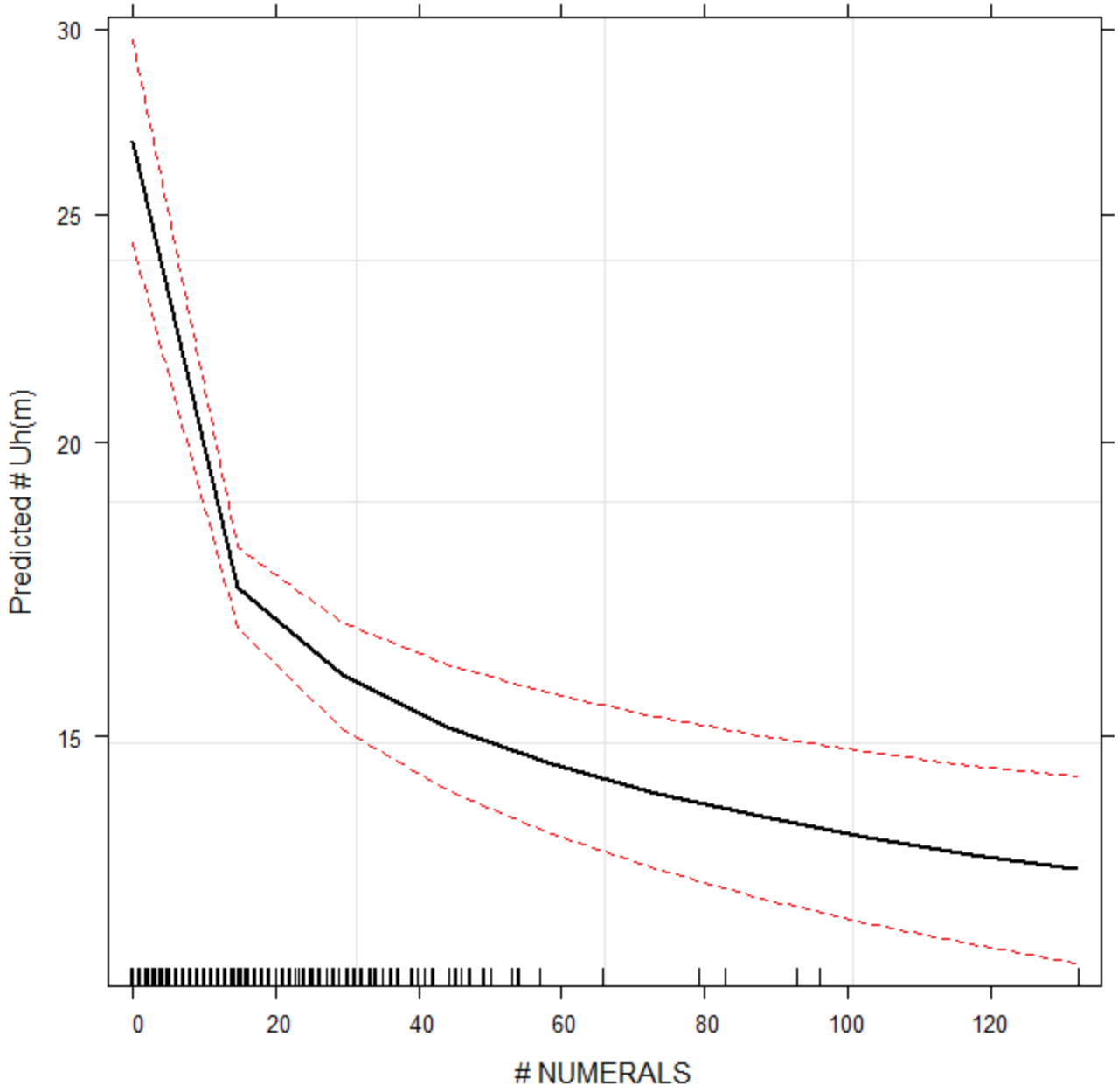
# Analysis

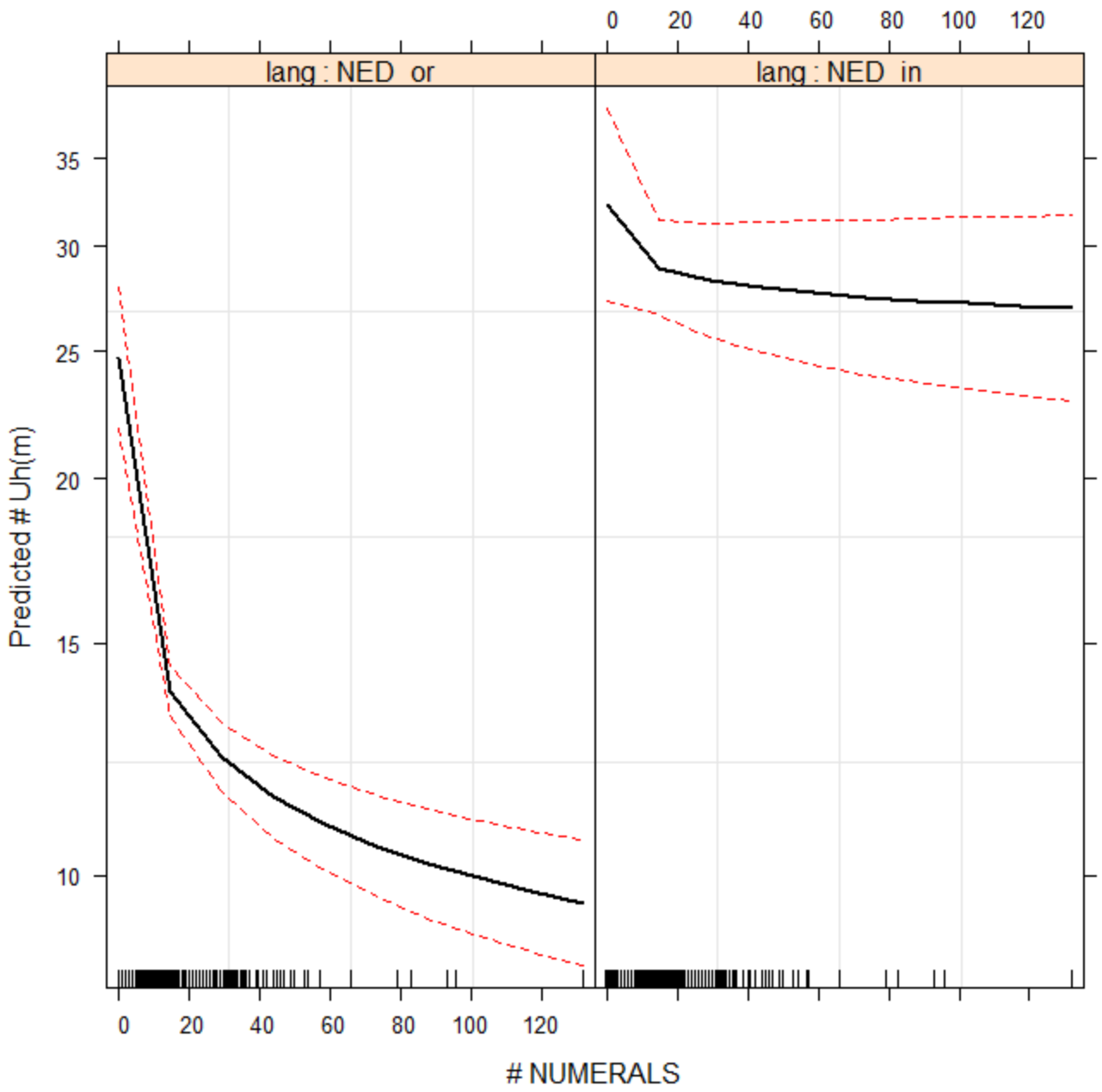
1. At the level of the sentences
2. To measure the effect of the French source load on the Dutch interpretations: At the level of the files

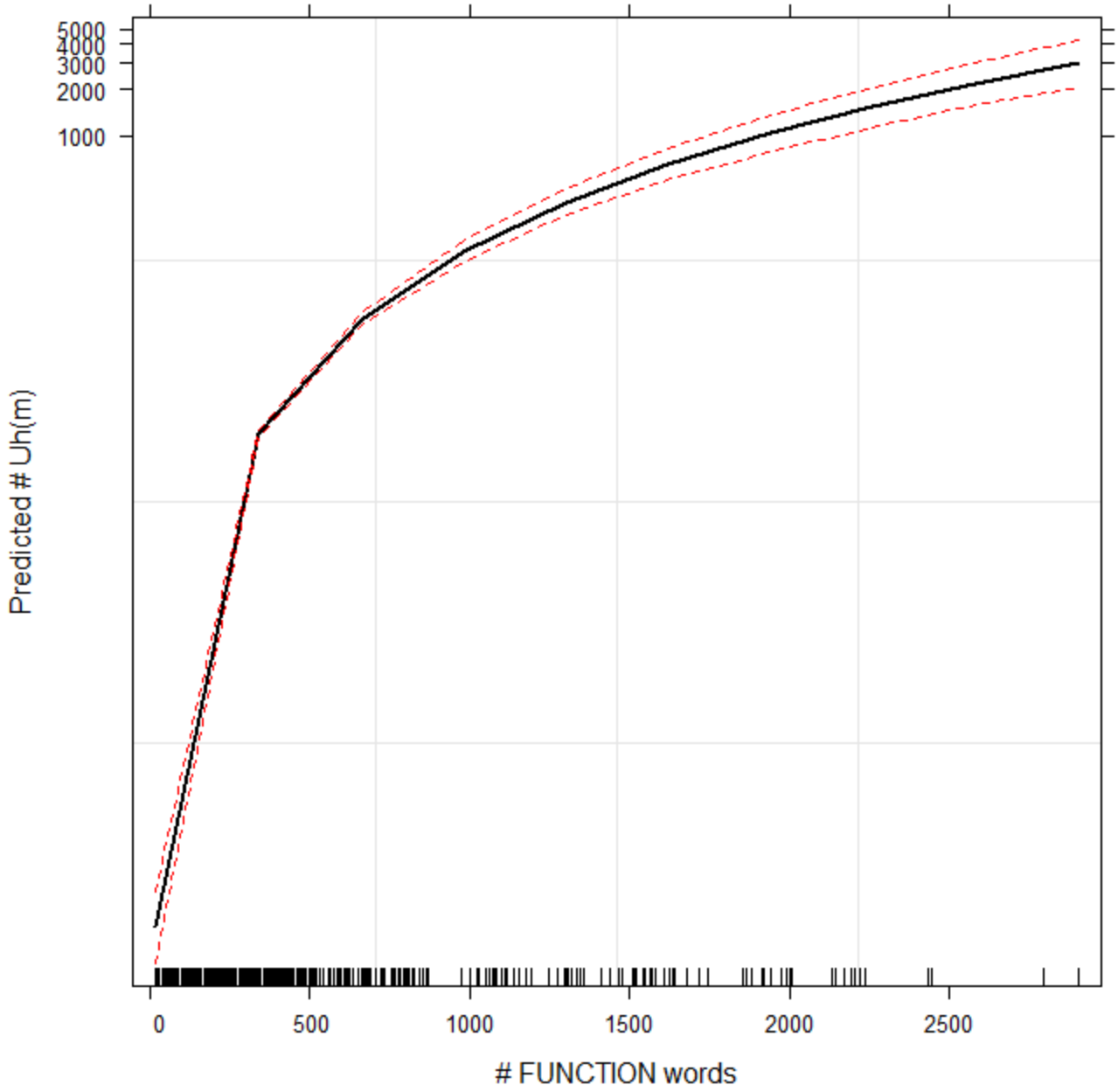

30
25
20
15
Predicted # Uh(m)
NED_or
NED_in
lang

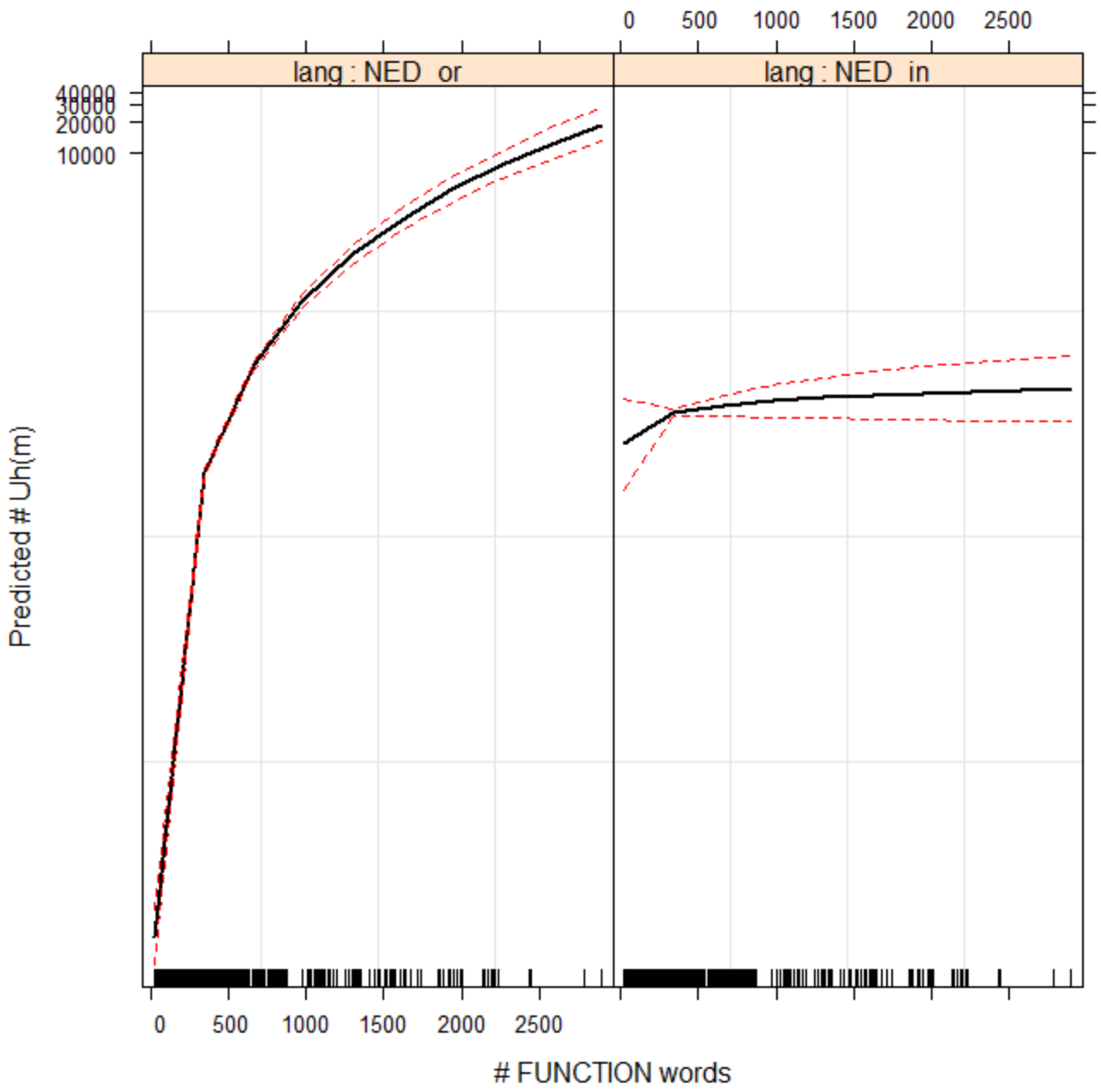

2000
1500
1000
500
Predicted # Uh(m)
0
500
1000
1500
# CONTENT words


lang : NED or
lang : NED in
Predicted # Uh(m)
# CONTENT words
25000
20000
15000
10000
5000
0
500
1000
1500


30
25
20
15
Predicted # Uh(m)
0
20
40
60
80
100
120
# NUMERALS


lang : NED or
lang : NED in
Predicted # Uh(m)
# NUMERALS
0
20
40
60
80
100
120
10
15
20
25
30
35


5000
4000
3000
2000
1000
Predicted # Uh(m)
0
500
1000
1500
2000
2500
# FUNCTION words


lang : NED or
lang : NED in
Predicted # Uh(m)
40000
30000
20000
10000
0
500
1000
1500
2000
2500
# FUNCTION words

# Conclusion

Confirmation:

- More *uh(m)* with interpreters than non-interpreters
- Lexical content has enhancing effect
- Numbers lead to high score of *uh(m)* throughout

Negative effect of lexical content for non-interpreters: scripted nature of parliamentary speeches

# Conclusion

Results demonstrate informational load:

Positive effect of grammatical material on the frequency of *uh(m)* for non-interpreters

Absent for interpreters

HENCE: Non-interpreters produce more *uh(m)* when they speak longer, interpreters when processing more content

# Conclusion

Future prospects:

- Syntactic embedding
- Position of *uh(m)* in utterance

Thank you!

koen.plevoets@ugent.be