A CROSS-LINGUISTIC STUDY OF THE RELATIONSHIP BETWEEN GRAMMAR & LEXICAL DEVELOPMENT

Maria Cristina Caselli

Antonella Devescovi

Daniela Marchione

Judy Reilly

Elizabeth Bates

IASCL-SRCLD 2002

**Thanks to DC01289, NS22343, DC00216, MacArthur Foundation*

PRIOR ITALIAN-ENGLISH COMPARISONS USING THE CDI

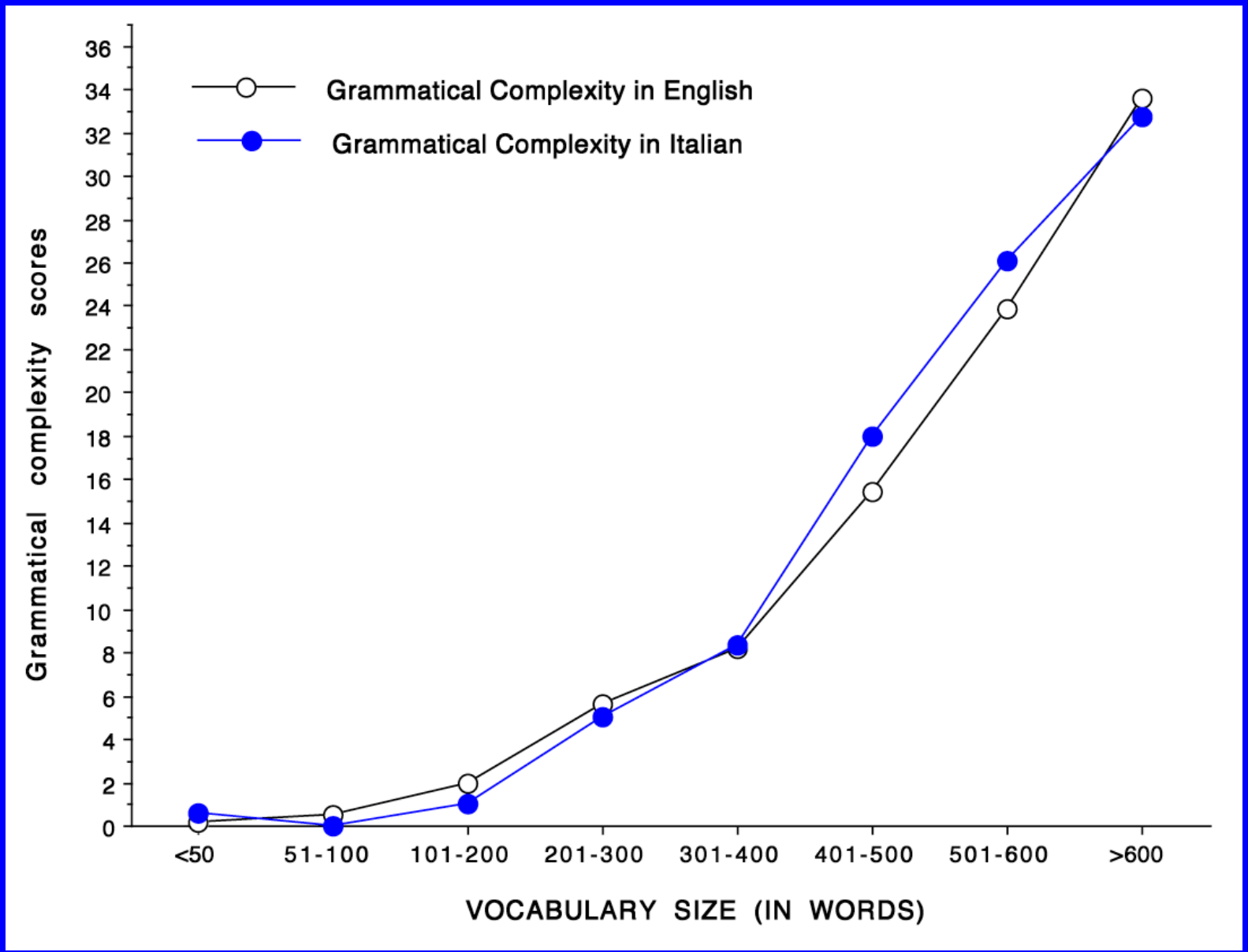
- Caselli, M. C., Bates, C., Casadio, P., Fenson, L., Fenson, J., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10,* 159-199.

- Caselli, M. C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language, 26,* 69-111.

[redaction of both papers published in M. Tomasello & E. Bates (Eds.), *Essential readings in language development.* Oxford: Basil Blackwell, 2001].

CROSS-LINGUISTIC SIMILARITIES IN EARLIER CDI STUDIES

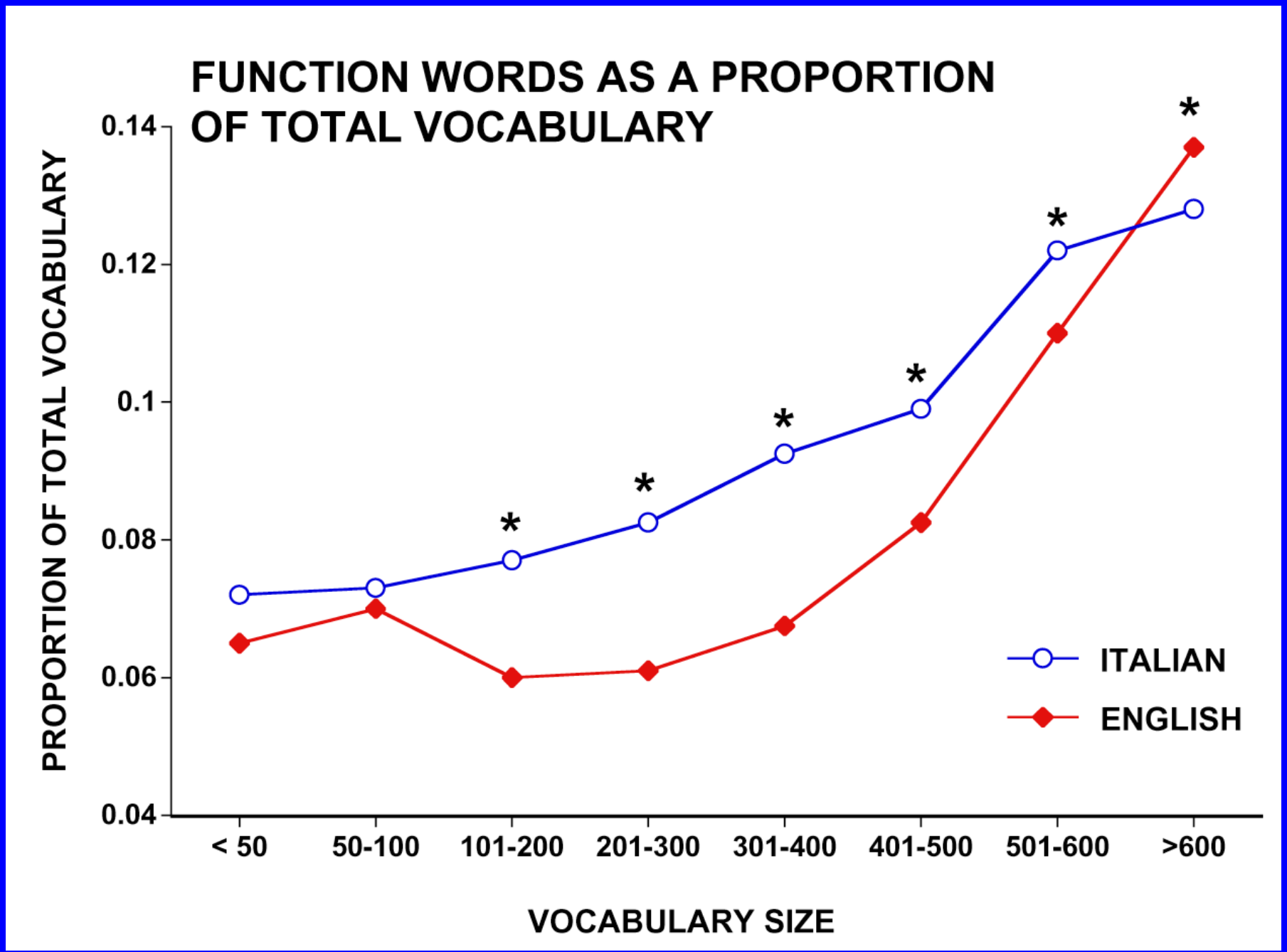
- Means and ranges in all scales of both lexical and grammatical development
- Same successive "waves" of lexical growth
  - Routines
  - Names for things
  - Verbs & adjectives
  - Grammatical function words
- Strong non-linear relationship between vocabulary size and grammatical complexity
  - Based on 37 sentence pairs selected to reflect contrasts known to develop from 16-30 months in each language


Grammatical Complexity in English
Grammatical Complexity in Italian
Grammatical complexity scores
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
<50
51-100
101-200
201-300
301-400
401-500
501-600
>600
VOCABULARY SIZE (IN WORDS)

CROSS-LINGUISTIC DIFFERENCES IN EARLIER CDI STUDIES

- Higher proportions of "social words" (routines, proper names) in Italian children from 8-30 months

- Differential growth trajectories for grammatical function words
  - Non-linear in English
  - Linear in Italian


FUNCTION WORDS AS A PROPORTION OF TOTAL VOCABULARY
PROPORTION OF TOTAL VOCABULARY
0.14
0.12
0.1
0.08
0.06
0.04
*
*
*
*
*
*
ITALIAN
ENGLISH
< 50
50-100
101-200
201-300
301-400
401-500
501-600
>600
VOCABULARY SIZE

CROSS-LINGUISTIC DIFFERENCES IN EARLIER CDI STUDIES

- Higher proportions of "social words" (routines, proper names) in Italian children from 8-30 months
- Differential growth trajectories for grammatical function words
  - Non-linear in English
  - Linear in Italian
- Informal examination of 3 longest utterances reported by parents suggested greater/earlier complexity in Italian children from 18-30 months

EXAMPLES OF SPEECH BY TWO-YEAR-OLDS IN DIFFERENT LANGUAGES
(underlining = content words)

English (30 months):

I	*wanna*	*help*	*wash*	*car*
1st pers. singular	modal indicative	infinitive	infinitive	

Italian (24 months):

Lavo	*mani,*	*sporche,*	*apri*	*acqua.*
Wash	hands	dirty	open	water
1st pers. singular indicative	3rd pers. feminine plural	feminine plural	2nd pers. singular imperative	3rd pers. singular

I wash hands, dirty, turn on water

QUESTIONS

- Can we capture cross-linguistic differences in grammatical development using parent report?
  - Three longest utterances reported by parents on the CDI Words & Phrases form
- Does the relationship between grammar & vocabulary differ over languages?
- What is the "right" coding scheme for cross-linguistic comparisons?
  - MLU in content words
  - MLU in total words (content + function)
  - MLU in morphemes (3 versions)

PARTICIPANTS

- Subsamples from national CDI norming studies
  - Same as Caselli et al., 1999
- 233 children in each language
  - Selected from > 1000 in English
  - Selected from > 600 in Italian
- Final subsamples matched for
  - Age
    - 18-30 months
  - Gender
    - 120 females, 113 males per language
  - Expressive vocabulary size
    - 50 - 680 words

MATERIALS

- Three longest utterances reported by parents
  - Eliminated all obvious cases of songs, prayers, counting & other formulae
  - MLU coded in five increasing coding schemes
    - Length in content words
    - Length in total words (content + function)
    - Length in Morphemes 1: conservative count
    - Length in Morphemes 2: expanded pronoun count
    - Length in Morphemes 3: expanded gender count
  - Averaged over utterances for each child, for each coding scheme

CODING FOR OBSERVED/ATTEMPTED

- All utterances coded in two forms
  - Observed: actual utterance reported
    - e.g. "Kitty sleeping"
  - Attempted: conservative expansion of reported utterance to restore grammaticality
    - e.g. "The kitty is sleeping"
  - Both observed and attempted utterances coded by native speakers, applying all five coding schemes
  - Ratios of observed/attempted reflects proportion of target utterances that children are able to produce in each language
    - e.g. "Kitty sleeping"/"The kitty is sleeping"
      - 2/4 words (50%) - 3/5 morphemes (60%)

MLU in MORPHEMES: THREE CODING SCHEMES

- MLU1: most conservative/traditional count
  - Markedness for nouns in both languages
    - Unmarked singular → no additional points
    - Marked plural → one additional point
  - Markedness for verbs in English
    - Unmarked zero form → no additional points
    - Marked all others → one additional point
  - Markedness for verbs in Italian
    - Unmarked 3rd pers sing → no additional points
    - Marked = all others → one additional point
  - Markedness for plural modifiers in Italian

MLU in MORPHEMES: THREE CODING SCHEMES

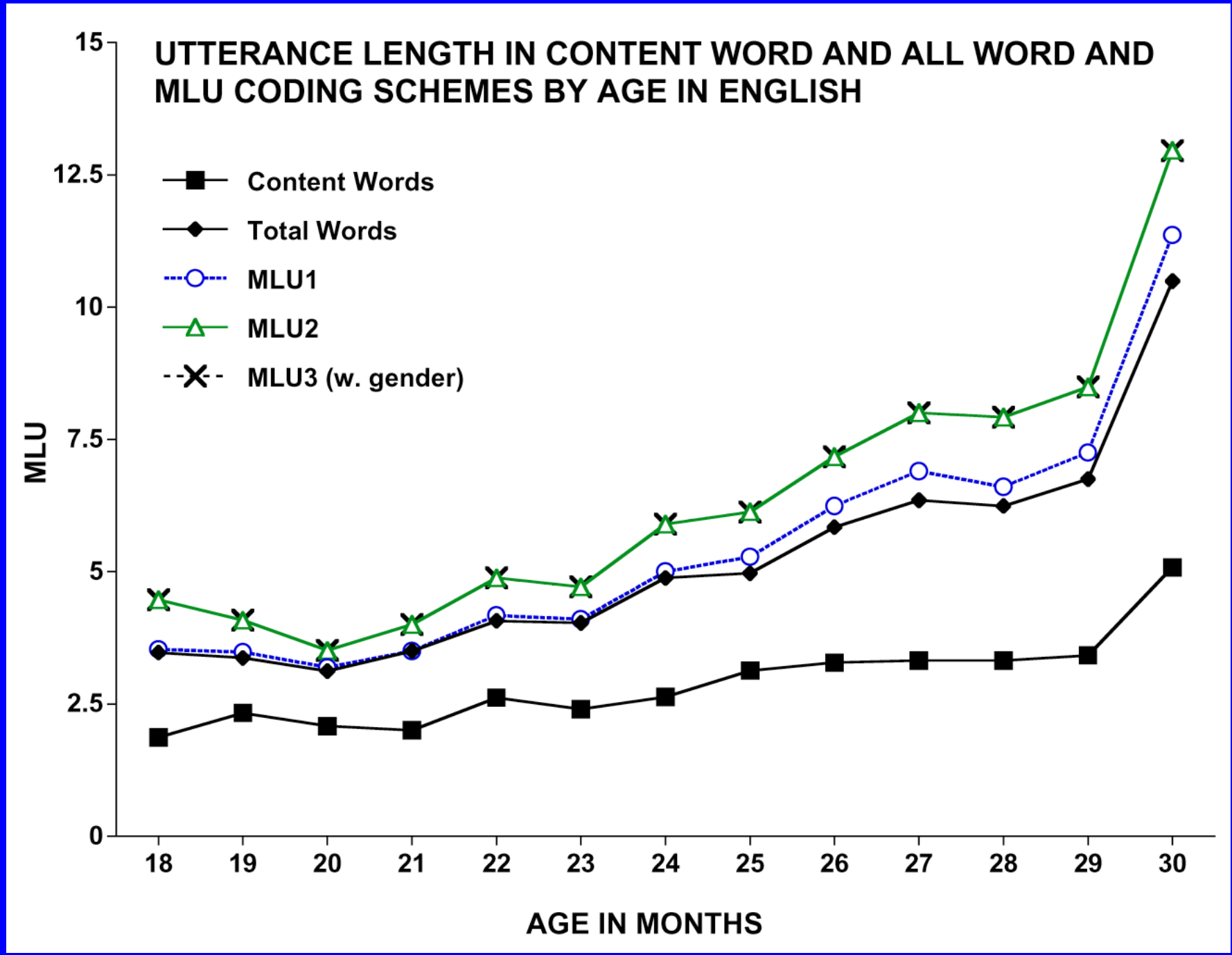
- MLU2: additional points for pronouns
  - Same assumptions in both languages
  - Unmarked = 3rd person singular subject pronoun
  - Marked: one point for each deviation
    - 1st or 2nd person → one additional point
    - Plural → one additional point
    - Object pronouns → one additional point
  - Pronominal modifiers treated like other modifiers in both languages
    - No additional points in English
    - Additional points for plural modifiers in Italian

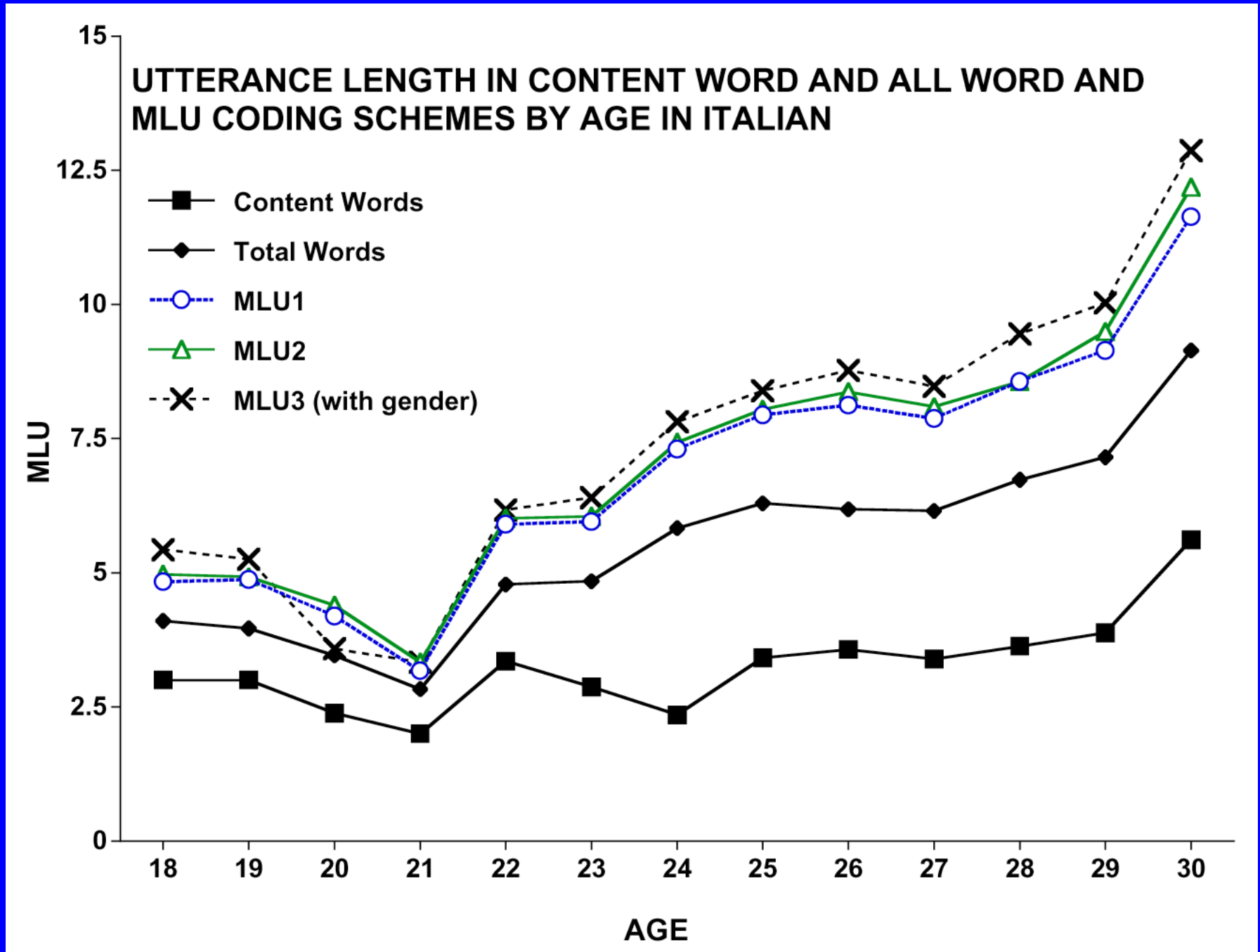
MLU in MORPHEMES: THREE CODING SCHEMES

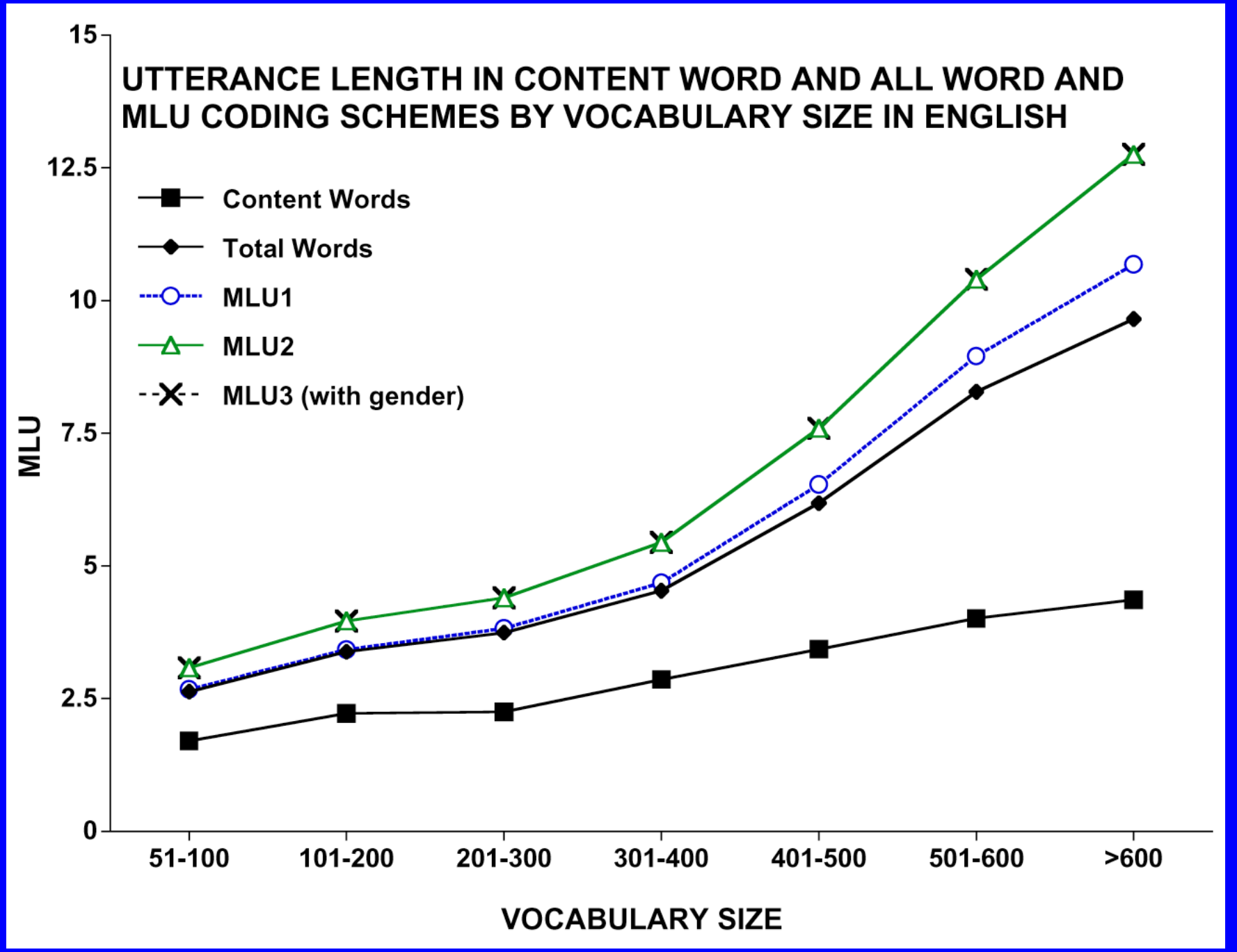
- MLU3: additional points for gender agreement
  - English
    - No additional points possible
    - MLU3 = MLU2
  - Italian
    - No additional points for gender on nouns or pronouns
      - Reject assumption that masculine = unmarked
    - Additional point for each gender-agreeing modifier
    - MLU3 > MLU2

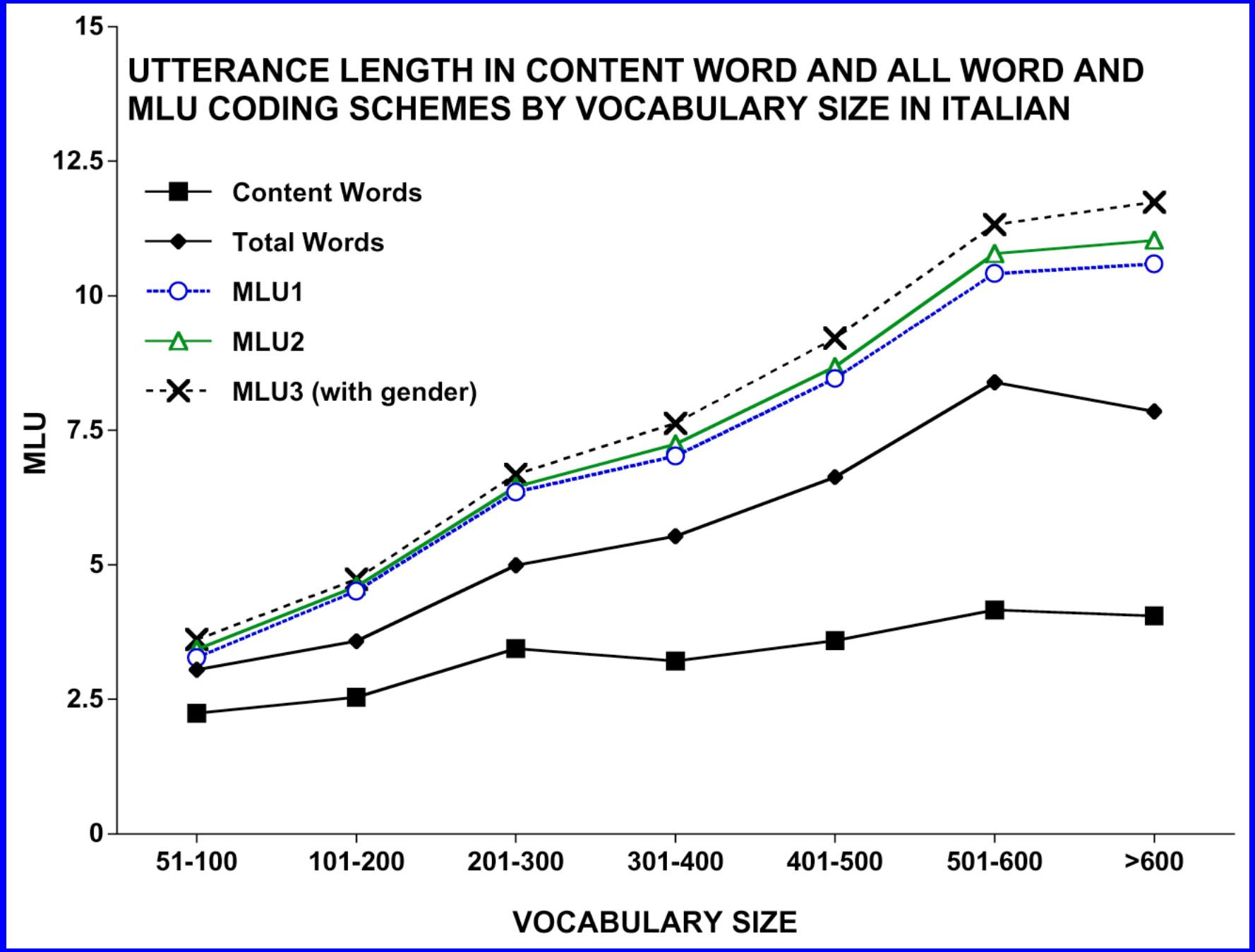
ANALYSES OF VARIANCE

- Developmental level analyzed two ways
  - Age: 18-30
  - Vocabulary Size
    - 50-100
    - 101-200
    - 201-300
    - 301-400
    - 401-500
    - 501-600
    - <600
- Language by Age or Voc Level by Coding Scheme


UTTERANCE LENGTH IN CONTENT WORD AND ALL WORD AND MLU CODING SCHEMES BY AGE IN ENGLISH
Content Words
Total Words
MLU1
MLU2
MLU3 (w. gender)
MLU
15
12.5
10
7.5
5
2.5
0
18
19
20
21
22
23
24
25
26
27
28
29
30
AGE IN MONTHS

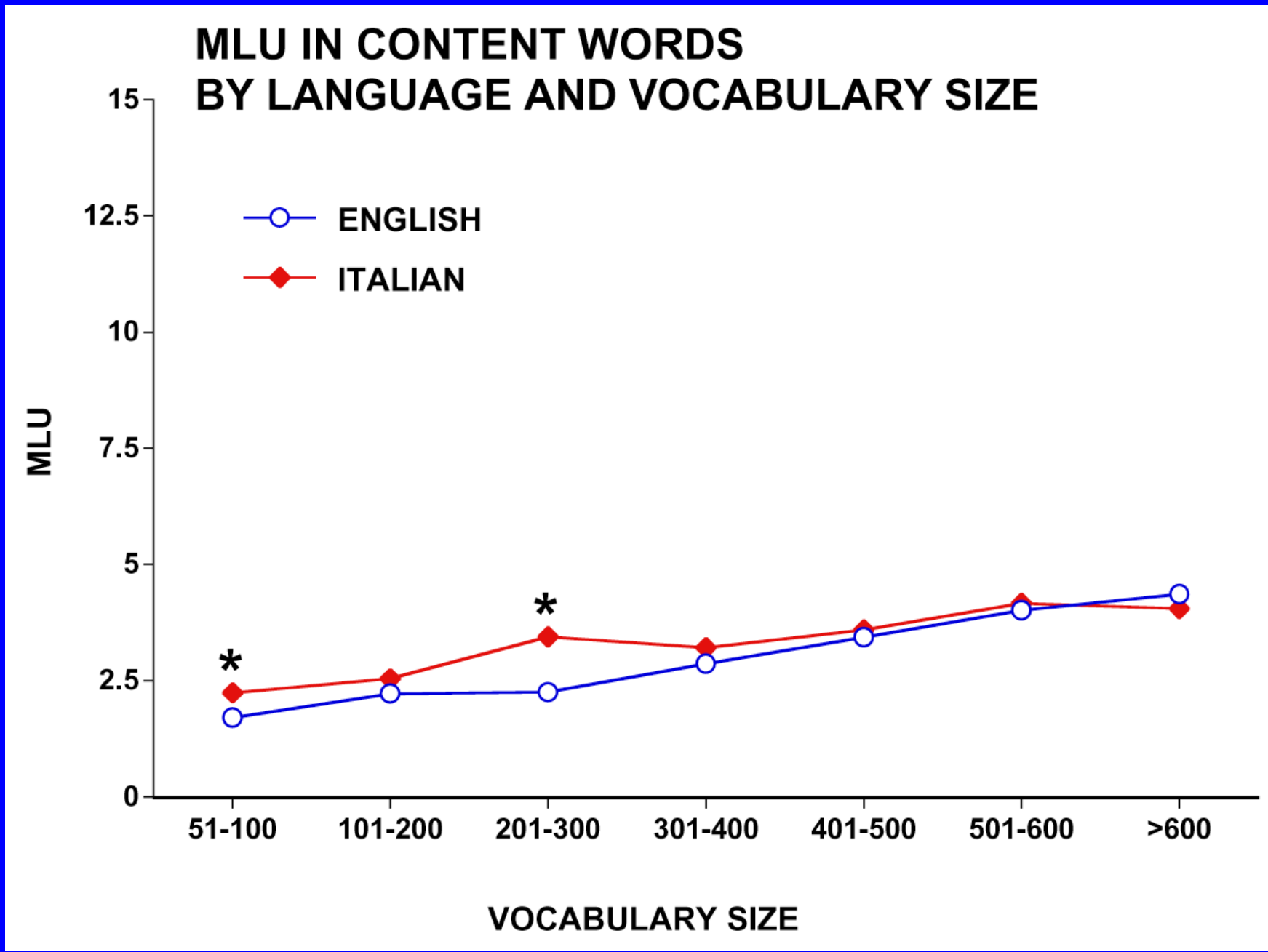

UTTERANCE LENGTH IN CONTENT WORD AND ALL WORD AND MLU CODING SCHEMES BY AGE IN ITALIAN
Content Words
Total Words
MLU1
MLU2
MLU3 (with gender)
MLU
AGE
0
2.5
5
7.5
10
12.5
15
18
19
20
21
22
23
24
25
26
27
28
29
30


UTTERANCE LENGTH IN CONTENT WORD AND ALL WORD AND MLU CODING SCHEMES BY VOCABULARY SIZE IN ENGLISH
Content Words
Total Words
MLU1
MLU2
MLU3 (with gender)
MLU
15
12.5
10
7.5
5
2.5
0
51-100
101-200
201-300
301-400
401-500
501-600
>600
VOCABULARY SIZE


UTTERANCE LENGTH IN CONTENT WORD AND ALL WORD AND MLU CODING SCHEMES BY VOCABULARY SIZE IN ITALIAN
Content Words
Total Words
MLU1
MLU2
MLU3 (with gender)
MLU
15
12.5
10
7.5
5
2.5
0
51-100
101-200
201-300
301-400
401-500
501-600
>600
VOCABULARY SIZE

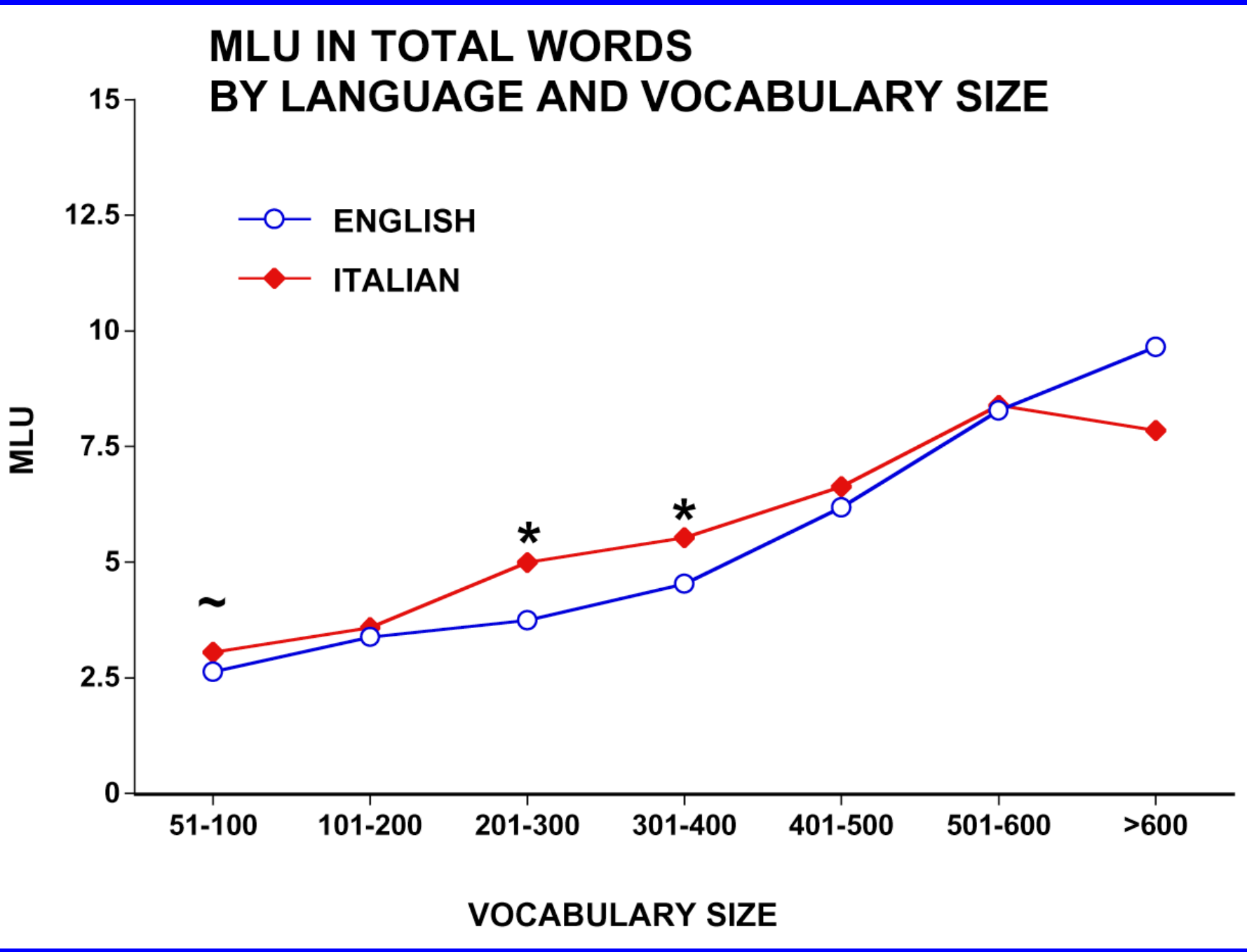
CROSS-LINGUISTIC DIFFERENCES: Interim Summary

- In magnitude
  - Italian > English
- In shape of growth
  - Italian = linear change over lexical levels
  - English = non-linear change over lexical levels
  - Similar to Caselli (1999) for closed-class proportion scores
- Vocabulary is a better predictor than age


MLU IN CONTENT WORDS
BY LANGUAGE AND VOCABULARY SIZE
ENGLISH
ITALIAN
MLU
15
12.5
10
7.5
5
2.5
0
*
*
51-100
101-200
201-300
301-400
401-500
501-600
>600
VOCABULARY SIZE

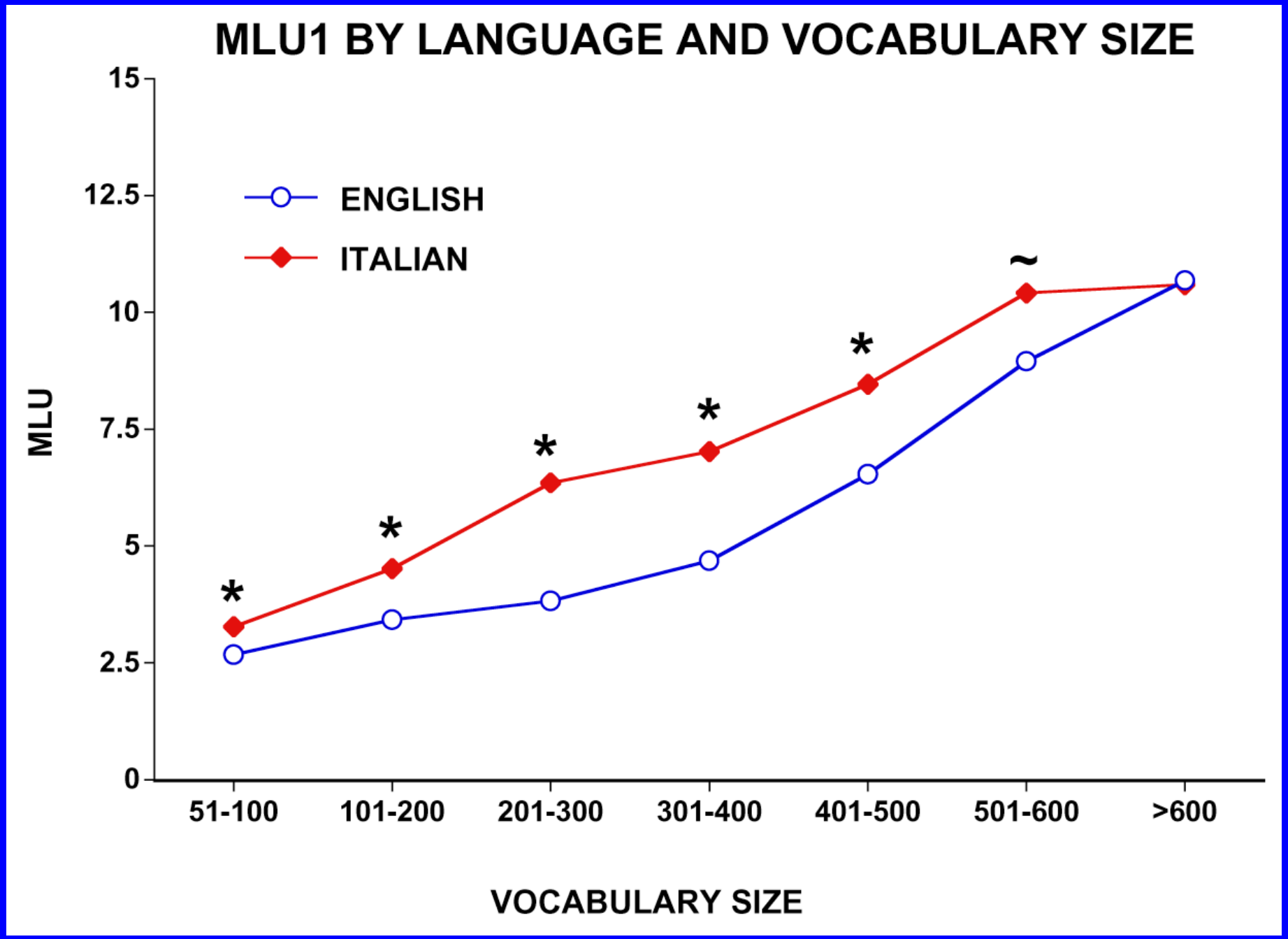
MLU IN CONTENT WORDS: TWO HYPOTHESES TO EXPLAIN THE ITALIAN ADVANTAGE

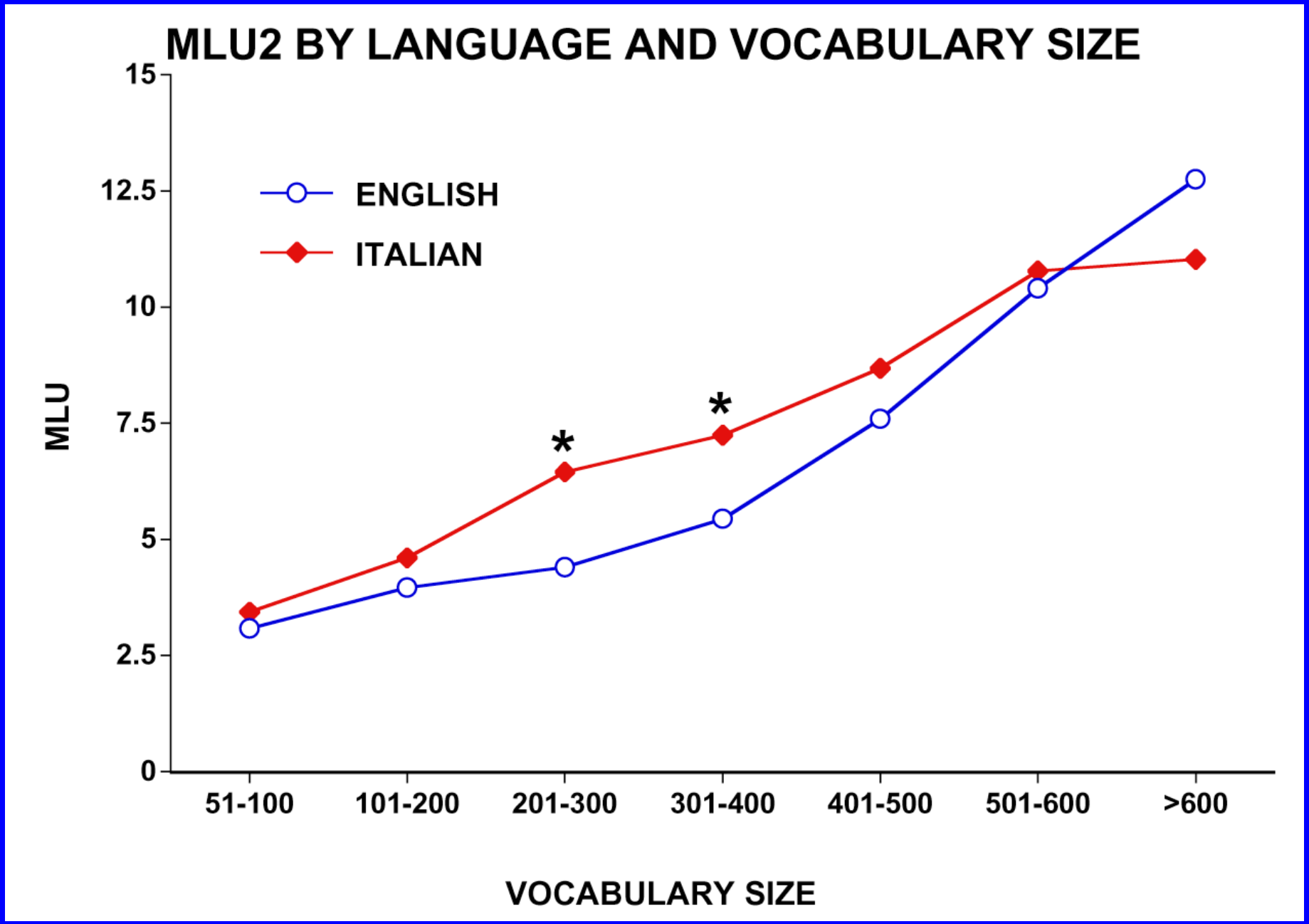
- By-product of social-word advantage
  - More proper-noun phrases in Italian?
  - NO: not verified by the data
- By-product of pro-drop
  - More pronominal subjects in English?
  - More nominal subjects in Italian?
  - YES: verified by the data
    - Only true in children < 300 words
    - Italian advantage disappears when proportion of nominal/pronominal subjects is controlled

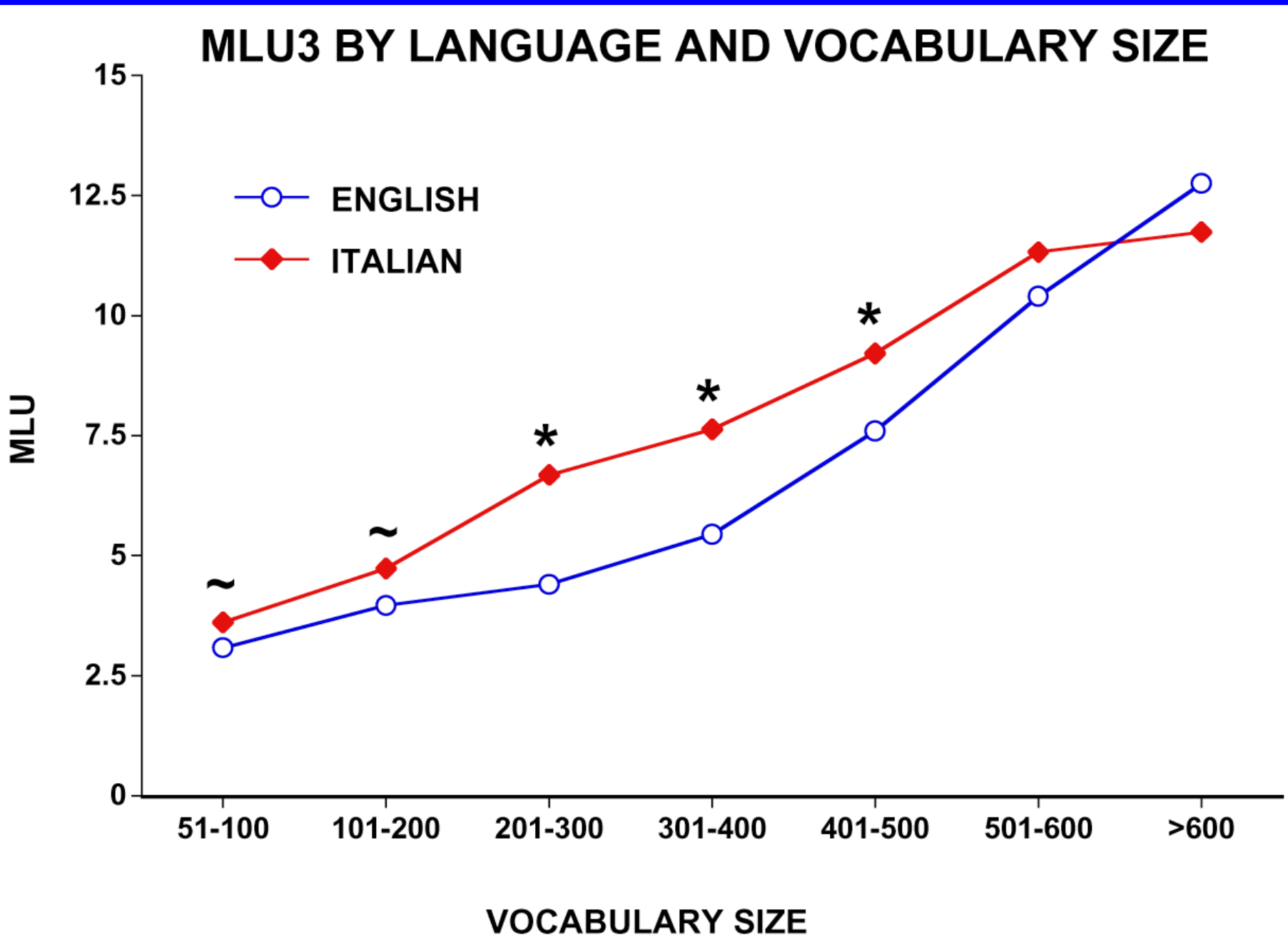

MLU IN TOTAL WORDS
BY LANGUAGE AND VOCABULARY SIZE
ENGLISH
ITALIAN
MLU
15
12.5
10
7.5
5
2.5
0
~
*
*
51-100
101-200
201-300
301-400
401-500
501-600
>600
VOCABULARY SIZE

MLU in WORDS: Interim Summary

- MLU in content words
  - Significant Italian advantage (< 300 words)
  - Not a "language-neutral" cross-language measure
  - Affected by grammar-specific properties
- MLU in total words
  - No significant Italian-English main effect
  - But this generalization may not hold up across all languages and/or discourse situations
  - Italian > English in number of contexts in which articles are obligatory (MacWhinney & Bates, 1978; Devescovi & Bates, 1989)


MLU1 BY LANGUAGE AND VOCABULARY SIZE
15
12.5
10
7.5
5
2.5
0
MLU
ENGLISH
ITALIAN
*
*
*
*
*
~
51-100
101-200
201-300
301-400
401-500
501-600
>600
VOCABULARY SIZE

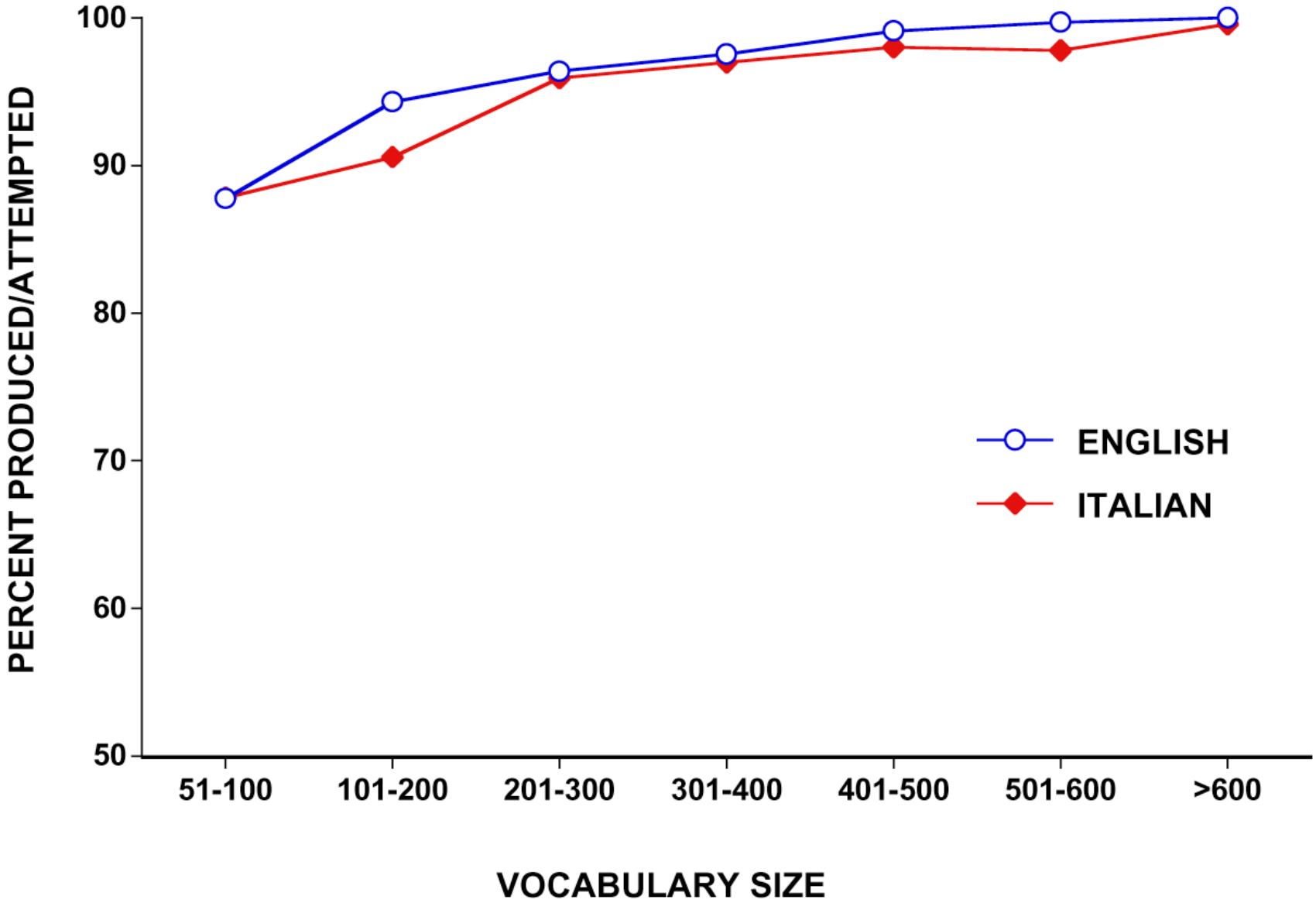

MLU2 BY LANGUAGE AND VOCABULARY SIZE
15
12.5
10
7.5
5
2.5
0
MLU
ENGLISH
ITALIAN
*
*
51-100
101-200
201-300
301-400
401-500
501-600
>600
VOCABULARY SIZE


MLU3 BY LANGUAGE AND VOCABULARY SIZE
15
12.5
10
7.5
5
2.5
0
MLU
ENGLISH
ITALIAN
~
~
*
*
*
51-100
101-200
201-300
301-400
401-500
501-600
>600
VOCABULARY SIZE

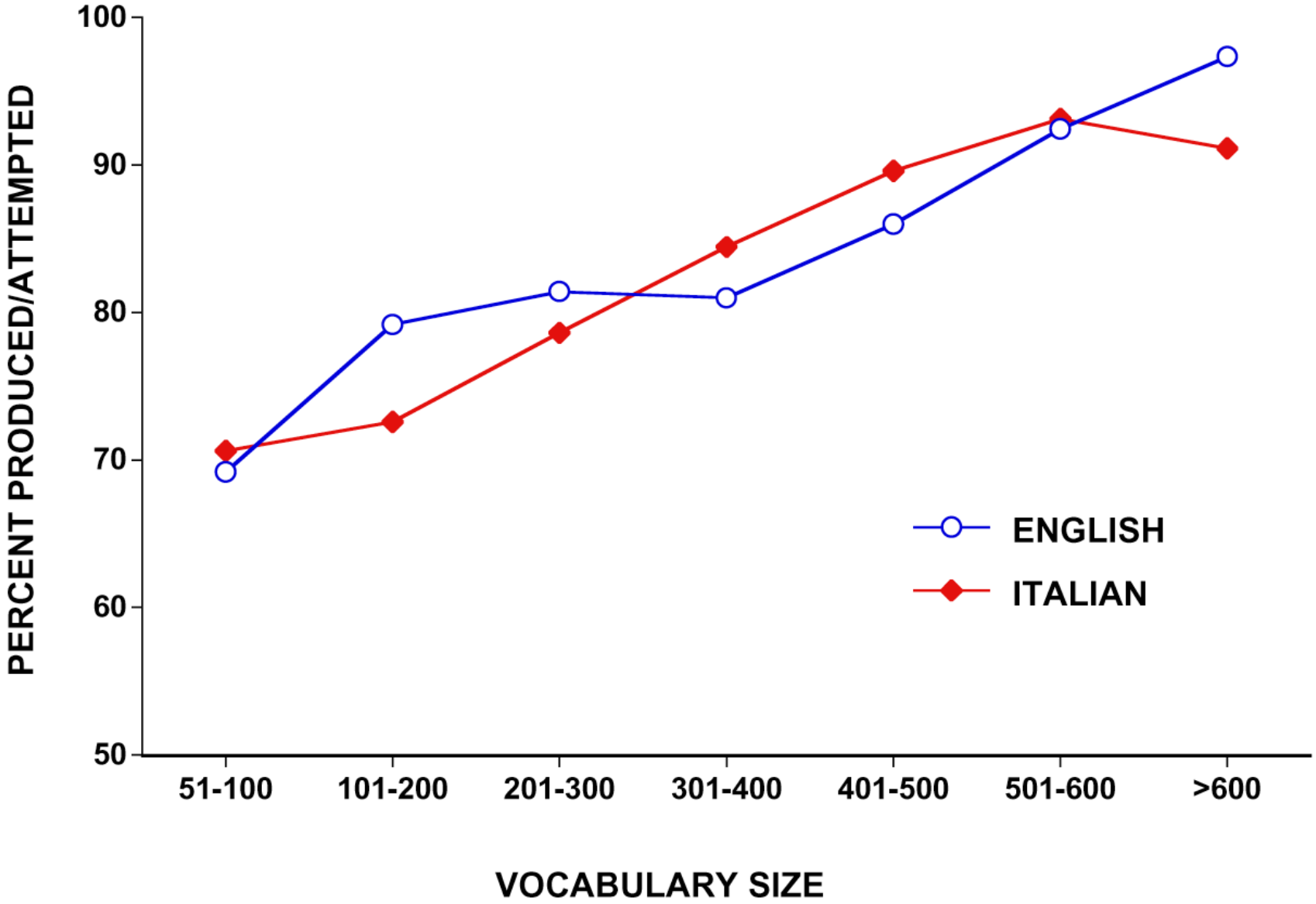
MLU in Morphemes: Interim Summary

- Italian > English in all three coding schemes
  - Most consistent in MLU1
  - English "closes the gap" in MLU2
    - "pronoun inflation"
- Both languages ceiling by >600 words
  - English children 'catch up' by creating "chains"
    - "We went to the zoo and saw an X, and a Y, and a Z…"
    - Results may differ for free-speech
    - Results may differ for average rather than longest utterances

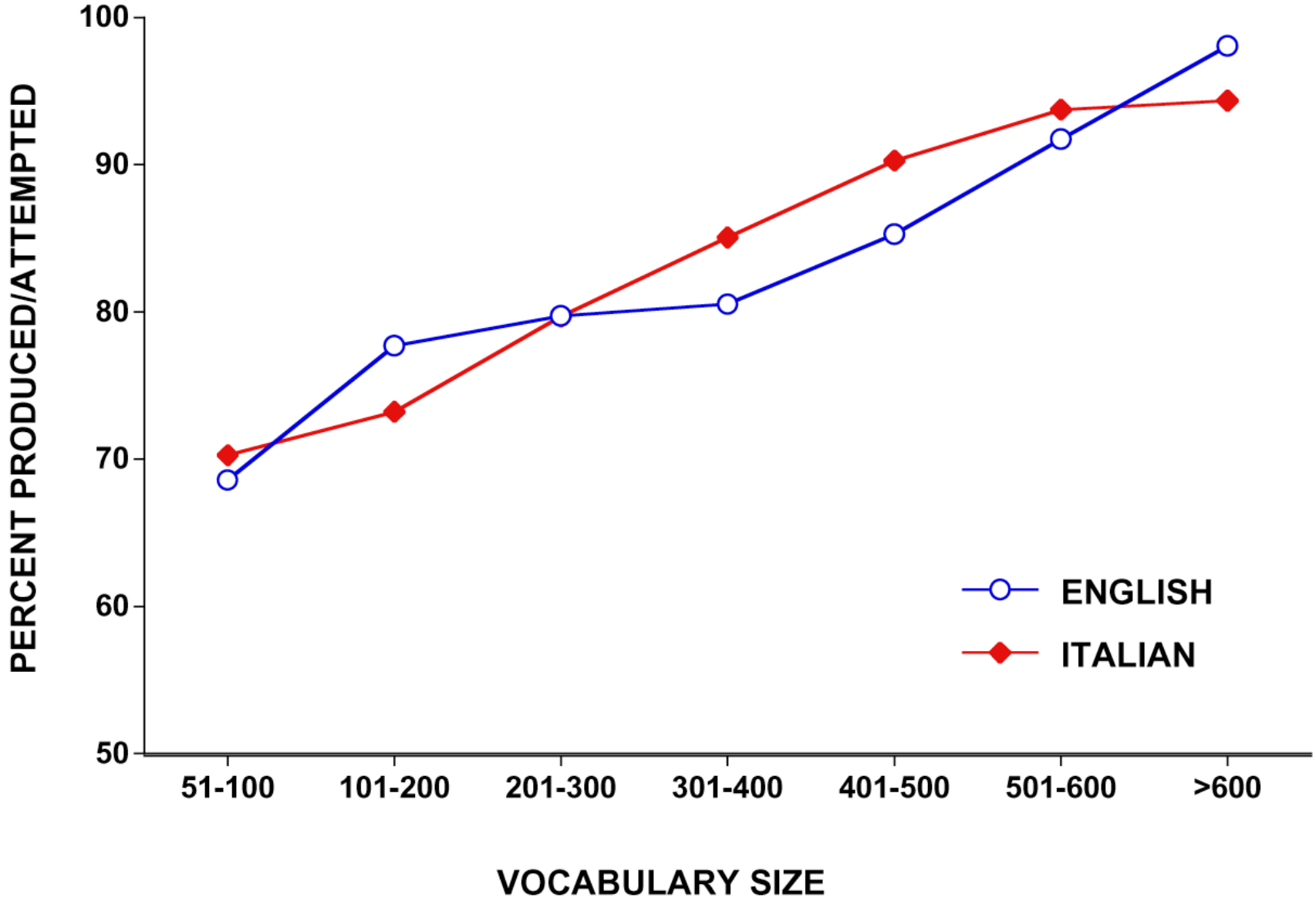
RATIO OF PRODUCED TO ATTEMPTED UTTERANCE LENGTH IN CONTENT WORDS



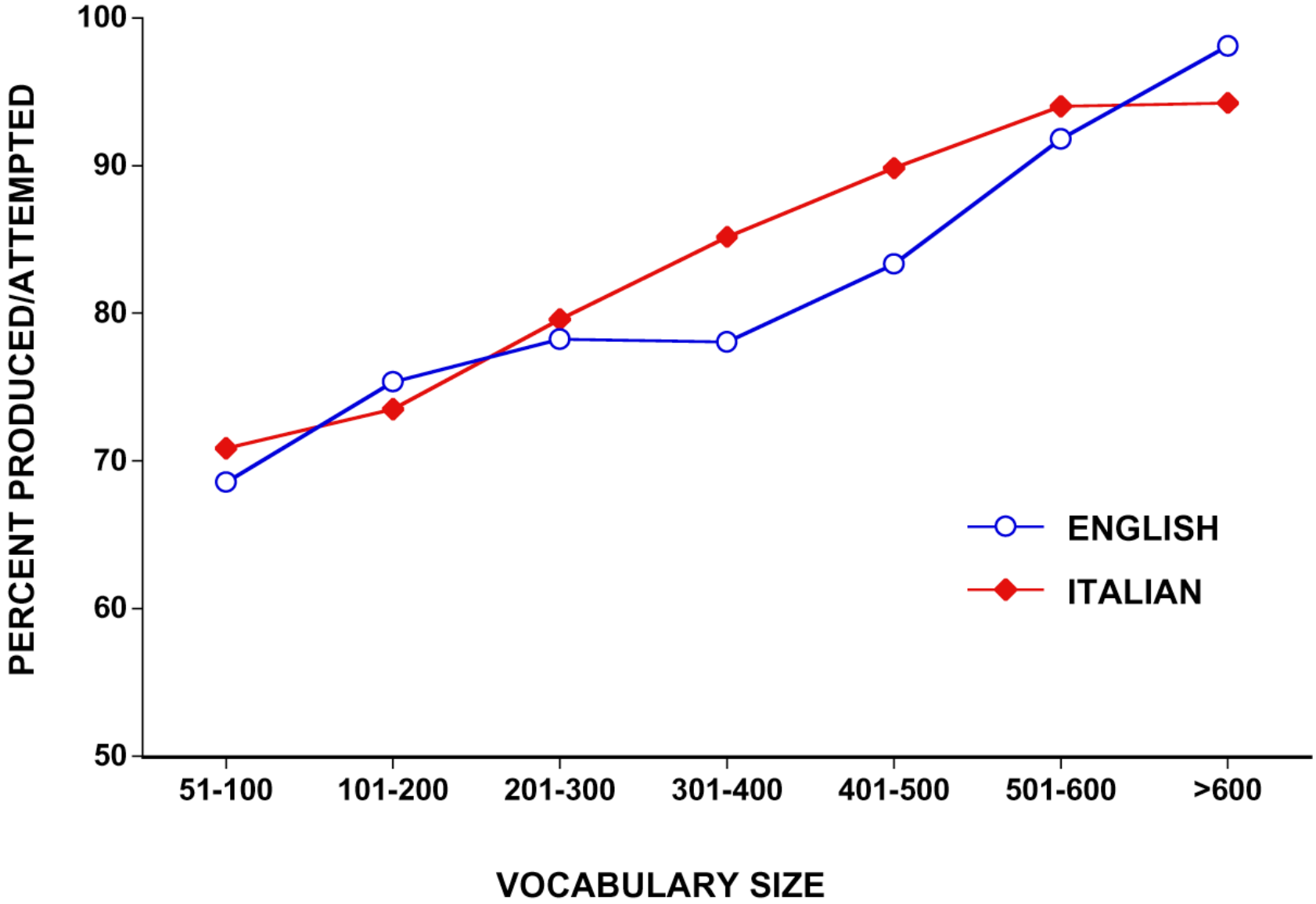
RATIO OF PRODUCED TO ATTEMPTED UTTERANCE LENGTH IN TOTAL WORDS BY LANGUAGE AND VOCABULARY SIZE



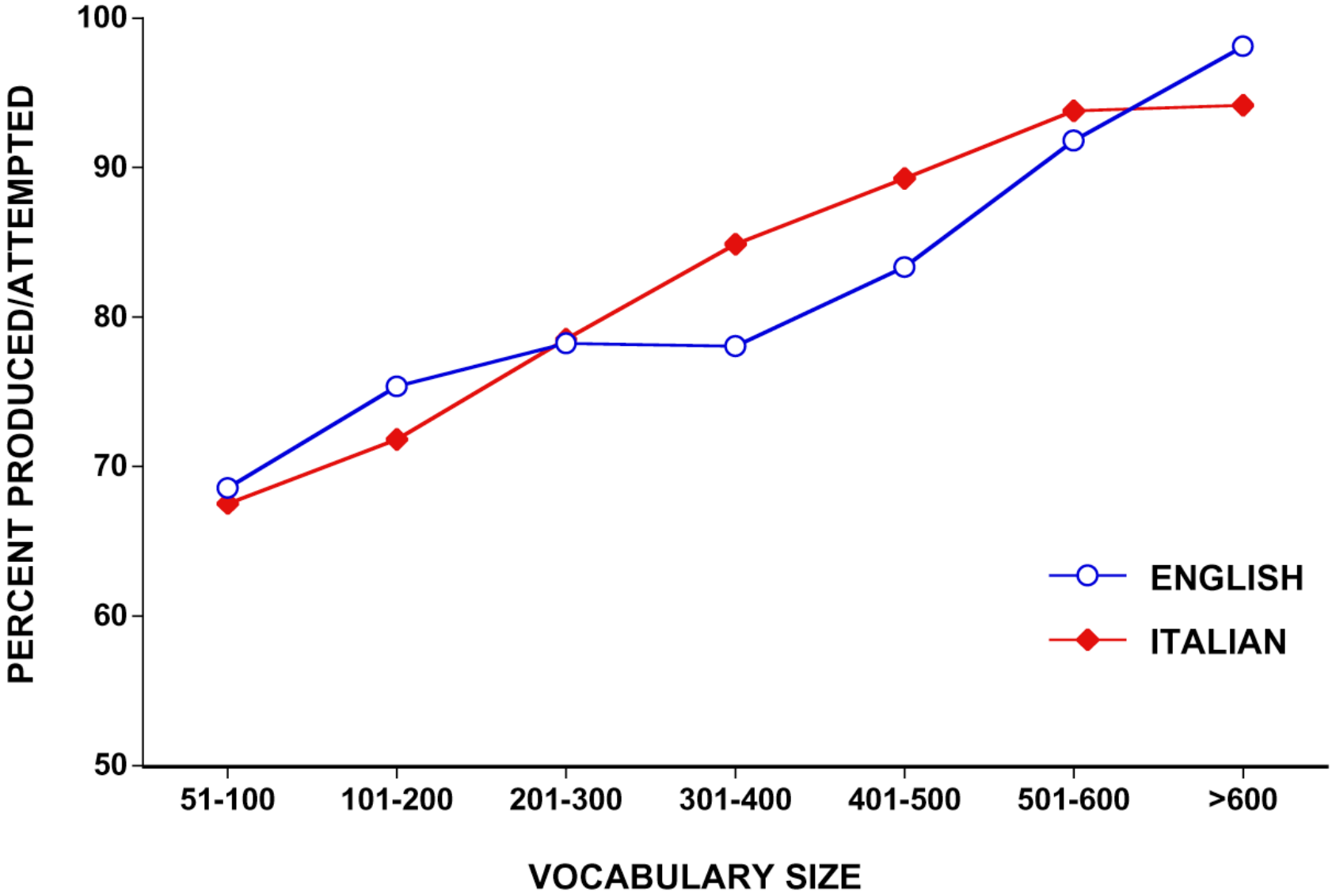
RATIO OF PRODUCED TO ATTEMPTED UTTERANCE LENGTH IN MLU1 BY LANGUAGE AND VOCABULARY SIZE



RATIO OF PRODUCED TO ATTEMPTED UTTERANCE LENGTH IN MLU2 BY LANGUAGE AND VOCABULARY SIZE



RATIO OF PRODUCED TO ATTEMPTED UTTERANCE LENGTH IN MLU3 BY LANGUAGE AND VOCABULARY SIZE



RATIO OF OBSERVED/ATTEMPTED: Interim Summary

- Significant increases by developmental level on all five coding schemes
  - "Closing the gap" between observed/attempted
  - "Zone of proximal development" for grammar?
- No language differences on any coding scheme
  - Compatible with conservative, input-driven models of grammatical development (e.g., Tomasello, Lieven, Pine)
  - Is there a universal constant in the grammatical "zone of proximal development"?

QUESTIONS ANSWERED

- Can we capture cross-linguistic differences in grammatical development using parent report?
  - YES
  - BUT ALL FINDINGS MUST BE REPLICATED
    - IN FREE SPEECH
    - IN AVERAGE VS. LONGEST UTTERANCES

QUESTIONS ANSWERED

- Can we capture cross-linguistic differences in grammatical development using parent report?
  - YES
- Does the relationship between grammar & vocabulary differ over languages?
  - YES
  - Linear in Italian, with earlier onset
  - Non-Linear in English, with initial delays

QUESTIONS ANSWERED

- Can we capture cross-linguistic differences in grammatical development using parent report?
  - YES
- Does the relationship between grammar & vocabulary differ over languages?
  - YES

- What is the "right" coding scheme for cross-linguistic comparisons?
  - No language-neutral or theory-neutral alternatives