Assessing morphological complexity: theoretical, methodological and educational issues

Gabriele Pallotti

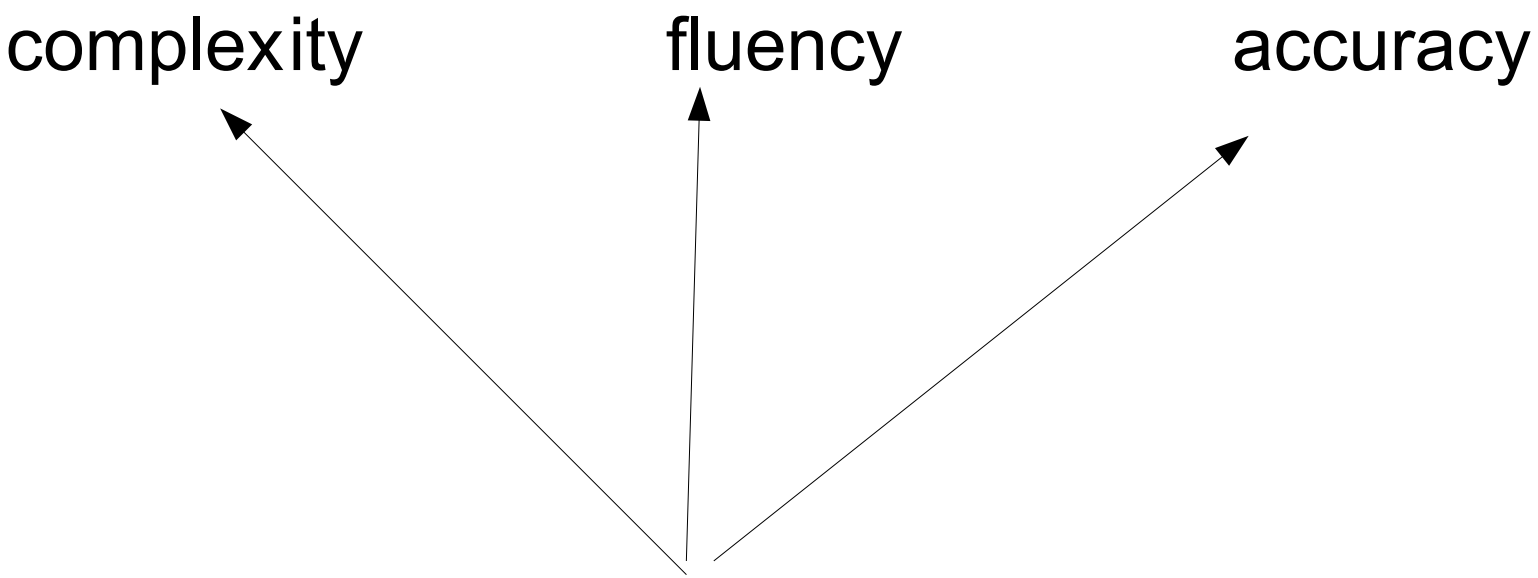
University of Modena and Reggio Emilia

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Summary

1. Linguistic complexity: definition, operationalization and practical applications
2. Morphological complexity in different languages, texts and learners
3. Complexity: ‘the more, the better?’
4. Implications for teaching and the concept of ‘appropriate complexity’

Complexity, accuracy, fluency (CAF)



Learning an additional language means building a more complex system, becoming more fluent and more accurate (Housen, Kuiken & Vedder 2012)

It also (or mainly) means using it more adequately! (Pallotti 2009, 2022; Kuiken & Vedder 2014; 2017, 2022; Kuiken, Vedder & Gilabert 2010)

Some common questions

- What is linguistic complexity? Can we measure it?
- Are some languages more complex than others?
- Are more complex languages more difficult to learn?
- Does a second or additional language grow more complex as one learns it?

Complexity as a key notion in SLA research

Three meanings of ‘complexity’

1. **Structural complexity**, a formal property of texts and linguistic systems having to do with the number of their elements and their relational patterns (= *complexity*)

2. **Cognitive complexity**, having to do with the processing costs associated with linguistic structures (= *difficulty*)

3. **Developmental complexity**, the order in which linguistic structures emerge and are mastered in second (and, possibly, first) language acquisition (= *development*)

Problems with polysemy

complex$_1$ structures are often more complex$_2$ and complex$_3$

=

complex structures are often more difficult and acquired late

this structure is complex$_3$ because it is complex$_1$ and complex$_2$

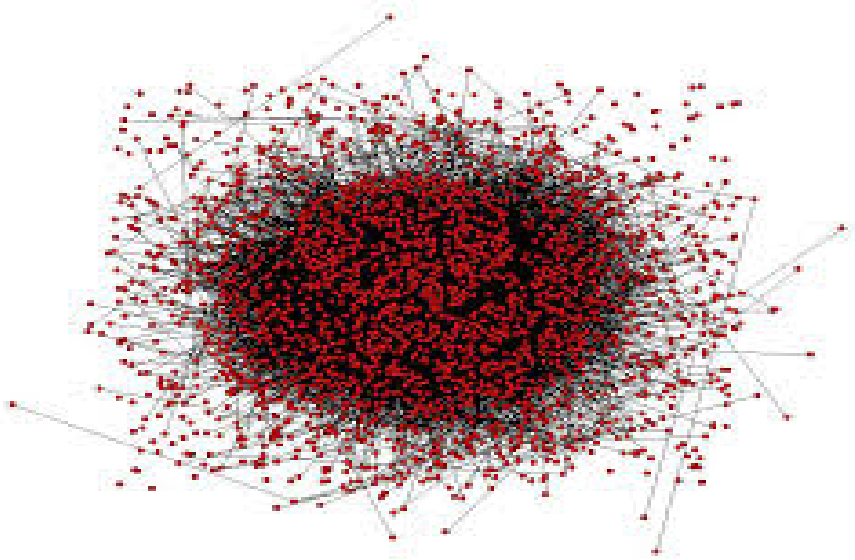
=

this structure is acquired late because it is complex and difficult

Structural complexity: a definition

The number of different elements and their interconnections in a text or a linguistic system, which both produce a longer description of the text's or system's structure

General definition of complexity: "a matter of the number and variety of an item's constituent elements and of the elaborateness of their interrelational structure, be it organizational or operational" (Rescher 1998:1)



UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

System vs Text complexity

System complexity = the complexity of a whole linguistic system, with all of its elements and rules (*langue,* competence; "theoretical complexity" Xanthos & Gillis 2010)

Text complexity = the complexity of a given piece of discourse (*parole,* performance; "observed complexity" Xanthos & Gillis 2010)

Morphological complexity (MC)

The complexity of inflectional processes in a linguistic system (e.g., English, Estonian, Italian) or in a text (letter, article, essay, conversation...).

Class	**Nominative** form: Final phoneme; example	**Genitive**: Formation from Nom.; example	**Partitive**: Formation from Nom.; example	Stem changes	Type %, CDS
I	V-final *kuu* 'moon'	Ø *kuu*	*+d* *kuu-d*	None	2.6
II	Ends in *-e* or *-s* *päike* 'sun'	*+(s)e* *päikese*	*+(s)t* *päikest*	If stem *-ne*, > *-se*	9.4
III	V-final *auto* 'car'	Ø *auto*	*+t* *auto-t*	None	7.6
IV	C-final *raamat* 'book'	+V *raamat-u*	+V*t* *raamat-ut*	+ V	12.5
V	V-final *maja* 'house'	Ø *maja*	Ø *maja*	None	9.5
VI	C-final *pilt* 'picture'	weak stem +V *pild-i*	strong stem +V *pilt-i*	Weakening gradation	49.2
VII	V-final *tigu* 'snail'	weak stem *teo*	Ø (strong stem) *tigu*	Weakening gradation	2.4
VIII	C-final *aken* 'window'	strong stem +V *akn-a*	weak stem +*t* *aken-t*	Strengthening gradation	2.8
IX	V-final *süda* 'heart'	+*me* *süda-me*	*+(n)t* *süda-nt*	Phonemic	1.0
X	C-final *kaas* 'lid'	*+(n)e* *kaa-ne*	*+(n)t* *kaa-nt*	Phonemic, +*e*	0.4
XI	V-final *lumi* 'snow'	*i>e* *lume*	*+(n)d* *lu-nd*	Vowel change	0.5
XII	V-final *jõgi* 'river'	weak stem, *i>e* *jõe*	strong stem, *i>e* *jõge*	Weakening gradation, *i>e*	0.3

(Vihman et al, 2021)

Potential applications of MC research

1. MC development over time in L1/L2 acquisition and language impairment

2. MC variation across tasks in L2 acquisition research

3. MC variation across genres: stylometry, comparative corpus linguistics

4. Assessing text difficulty

Previous definitions of morphological complexity in texts

In L2 acquisition studies:

Frequency of tensed forms, Number of different verb forms, Variety of past tense forms (Bulté and Housen 2012)

In L1 acquisition studies:

Inflectional Diversity (ID; Richards & Malvern 2004)

(Normalized) Mean Size of Paradigm (NMSP; Xanthos & Gillis 2010)

Analogy with lexical complexity

Complexity = high diversity of types with low repetition of tokens

Lexical complexity

talk, write, drink **>** talk, talk, talk (or talk, talking, talks)

Morphological complexity

talk, talking, talks **>** talking, talking, talking (or talking, writing and drinking).

Standardized Type/Token Ratio (STTR)

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Inflectional diversity (Richards & Malvern 2004)

1) Calculate D (lexical diversity) for inflected word forms (*go, going, goes, went* = 4 types) and for lemmas (*go, going, goes, went* = 1 type).

2) Inflectional diversity (ID) = $D_{\text{infl w forms}} - D_{\text{lemmas}}$

Problems

- ID decreases as D increases: in a sample containing unique lemmas only (lexical TTR = 1), ID = 0 (# of word forms = # lemmas)

- ID is sensitive to sample size with samples < 200 words (Xanthos & Gillis 2010).

(Normalized) mean size of paradigm
(Xanthos & Gillis 2010)

1) From the entire corpus, draw subsamples of 50 words

2) For each subsample, and for each word class, calculate the number of inflected forms (size of paradigm for that subsample)

3) Average these values to obtain a mean size of paradigm

Problems

- different densities of a given word class in 50-word subsamples may produce +/- large paradigms

- relatively insensitive to text size, but highly sensitive to subsample size (MSP 500 > MSP 50)

A simple approach to calculating a text's morphological complexity

(Pallotti, 2015)

1. LINGUISTIC ANALYSIS

- Compute the number of different exponences (inflectional forms) in subsamples of N tokens of a given word class (e.g. nouns, verbs etc)

2. MATHEMATICAL ANALYSIS

- Compute variety within and across subsamples

1. Linguistic analysis

Inflectional processes

base + exponence (process) = inflected word **form**

book → *book-s* (concatenative process)

buch → *bücher : buch* + *er* + *umlaut* (concatenative and non-concatenative process)

kitab → *kutub* (non-concatenative process)

Describing inflections

	DB	sample IW(s)	exponence
WF is identical to DB	cut	cut (present or past tense)	Ø
WF consists in DB + additional graphemes at the end of the DB	cut rise, take talk	cuts risen, taken talked	s n ed
WF consists in DB minus some graphological material at the end of the DB	hide	hid	£e
WF consists in DB minus some graphological material in the middle of the DB	feed lead	fed led	_£e_ _£a_
WF consists in DB + additional graphemes replacing parts of the DB at the end of the DB	buy think	bought thought	uy/ought ink/ought
WF consists in DB + additional graphemes replacing parts of the DB in the middle of the DB	find, grind drive, ride	found, ground, bound drove, rode	_i/ou_ _i/o_
multiple aspects	keep, feel break, steal swear, tear	kept, felt broke, stole sworn, torn	_£e_t _ea/o_e _ea/o_n

Diversity of form-function relationships?

Same procedure, but instead of counting exponents (forms), count form-function relationships.

This can be operationalized by looking at strings encoding forms and functions as in standard morphemic transcriptions, e.g.:

- German: *en:*1pl.prs.ind; *en:*3pl.prs.ind; *en*:inf

- Italian: *i*:2sg.prs.ind; *i*:3sg.prs.sbjv

Problems

- what functional features are to be encoded? E.g. shuld one encode just 'present' or 'present, habitual, indicative'?

- how can one be sure of the functions of grammatical forms in an interlanguage? E.g. does -*ing* correspond to present, progressive, indicative, or just present or just progressive?

Interlanguage morphology

Relatively easy cases:

they find-s

they find-ed

she find- Ø

asess-ed

impast-ed

More complex cases:

'commite' = commit-e or *commite-Ø?*

2. Mathematical analysis

Computing morphological complexity (MC 10)

For each word-class (e.g. nouns, verbs, adjectives) create sets of N (e.g 10) tokens

For each set, count the exponents' types (min 1 – max 10); then compute the average set-internal variety. **(6+7)/2 = 6.5**

For each set pair, count exponents that are not shared (min 0 – max 20); then compute the average between-set diversity and divide it by two. **5/2 = 2.5**

Add the set-internal diversity score to the between-set diversity score/2, then subtract 1, to arrive at a global inflectional diversity score (morphological complexity).

6.5 + 2.5 - 1 = 8.0 (MC10)

ed	ed
ed	**came**
took	**went**
was	was
Ø	Ø
Ø	Ø
ing	Ø
are	Ø
are	are
are	**is**
6	7
took, ing, came, went, is = 5	

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

General formula

MC = (within-subset variety + between-subset diversity/2) - 1

Value range for MC10

Min 0 = 1+0-1 / max 19 = 10+(20/2)-1

Analyse your texts with *Morpho complexity tool*

Alpha version

PLEASE NOTE: This tool is still under development and is not intended for general use yet. Current major limitations: 1) Analysis of nouns and periphrastic morphemes hasn't been systematically implemented; 2) analysis for English is still very preliminary and that for Italian is based on theoretical models which will be revised soon; 3) analysis for German, French and Spanish hasn't been implemented yet.

The mathematical computation of the Morphological Complexity Index (MCI) can be considered more stable and can be used by inputting any list of morphological exponences in the boxes in the following page and then hitting the 'calculate MCI' button.

1. Paste the text you want to analyse into the text box below. 500000 characters left.

2. Select language: English

3. Choose settings options: ☑ exclude proper nouns ☐ identify periphrastic morphemes

Analyse text now Clear

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

VERBS

All verb exponences in text:

Ø, ed, had, Ø, ed, was, Ø, Ø, ed, was, _e/ten, 's, was, ed, Ø, Ø, Ø, Ø, ed, Ø, Ø, Ø, Ø, Ø, s, were, ing, ed, had, been, ed, Ø, had, ed, was, ed, ing, ed, sought, Ø, had, been, ed, were, ing, is, n, ed, Ø, Ø, were, ed, ed, Ø, had, ed, ed, were, ing, is, Ø, were, were, ing, ed, Ø, 's, Ø, was, ed, was, ed, was, was, ing, are, Ø, Ø, Ø, is, _e/ten, 's, s, Ø, Ø, Ø, Ø, is, Ø, Ø, Ø, Ø, Ø, Ø, Ø, Ø, Ø, ing, Ø, Ø, Ø, ed, Ø, Ø, Ø, is, are, ing, Ø, Ø, are, ed, was, Ø, was, Ø, is, ing, ne, were, ing, ed, Ø, Ø, Ø, Ø, ed, was, ing, has, Ø, ing, Ø, Ø, Ø, 're, is, _k/d_, ed, is, Ø, is, Ø, went, Ø, Ø, Ø, Ø

NOUNS

All noun exponences in text:

s, Ø, Ø, s, Ø, Ø, Ø, Ø, Ø, s, s, s, Ø, Ø, s, Ø, Ø, Ø, s, Ø, Ø, s, s, Ø, Ø, s, s, s, s, Ø, Ø, s, Ø, Ø, Ø, Ø, Ø, Ø, Ø, s, s, s, Ø, s, s, s, s, s, Ø, s, Ø, s, Ø, Ø, Ø, Ø, Ø, s, Ø, Ø, s, Ø, s, Ø, Ø, Ø, Ø, s, s, s, Ø, s, Ø, Ø, Ø, s, s, Ø, Ø, Ø, s, s, s, Ø, s, s, Ø, Ø, Ø, s, Ø, Ø, Ø, Ø, Ø, Ø, Ø, s, Ø, Ø, s, Ø, Ø, s, Ø, s, Ø, Ø, Ø, Ø, Ø, Ø, s, Ø, Ø, Ø, Ø, Ø, Ø, Ø, s, s, s, Ø, Ø

PARAMETERS

segment size 10 random trials 100

Calculate MCI

RESULTS

Morphological complexity - VERBS: (5.57 + 4.97/2) - 1 = 7.05

Morphological complexity - VERBS (100 randomised trials): (5.43 + 5.16/2) - 1 = 7.01

Morphological complexity - NOUNS: (1.93 + 0.14/2) - 1 = 1

Morphological complexity - NOUNS (100 randomised trials): (1.99 + 0.02/2) - 1 = 1

5164	the	DT	the	
5165	United	NP	United	
5166	States	NPS	States	
5167	was	VBD	be	was
5168	n't	RB	n't	
5169	going	VBG	go	ing
5170	to	TO	to	
5171	collapse	VB	collapse	Ø
5172	if	IN	if	
5173	Chile	NP	Chile	
5174	became	VBD	become	_o/a_
5175	independent	JJ	independent	
5176	.	SENT	.	
5177	[S_END]			
5178	[S_START]			
5179	Why	WRB	why	
5180	were	VBD	be	were
5181	we	PP	we	
5182	so	RB	so	
5183	concerned	JJ	concerned	
5184	about	IN	about	
5185	it	PP	it	
5186	?	SENT	?	

MCI = 4.1

```
ing, ed, is, ing, is, ∅, ∅, ∅, ∅, ∅, _k/d_, ed, ing, ing, is, ∅, is, ∅, ∅, ∅,
∅, ∅, ∅, ∅, s, ing, ∅, ∅, ∅, ∅, ing, ing, ∅, ∅, ∅, are, ing, ed, ∅, ing, are,
∅, ing, ed, ∅, ing, ed, ∅, ∅, ed, ed, is, ∅, ∅, ing, ∅, ∅, ed, ed, ∅, ∅
```

VERBS AND NOUNS IN TEXT

In my **opinion**, **saying** that in a **world** **dominated** by **science** **technology** and **industrialisation**, there **is** no longer a **place** for **dreaming** and **imagination** **is** false . **Imagination** and **dreams** **belong** to **mankind** and **people** **use** the technological **progress** to **live** **better** their **life** and to **realise** their **dreams**. Most of the scientific and medical **researches**, **made** by the **use** of the **science** **technology**, **realised** the **dream** of **living** in a better **world**: **today**, **using** a personal **computer**, it **is** possible to **discover** the **cause** of a **disease** and , by its **analysis**, it **is** possible to **find** its **cure** and to **save** many human **lives**. **Industralisation** and **science** **technology** after **people** the **tools** to **communicate** as quick as possible even though they **live** away from each other . The **telecomunication** **system** **reduce** **time** and **distances** among **people** of different **countries**. The **use** of an Internet **program**, for **instance**, **make** you **talk** to an unknown **citizen** who **lives** on the other **side** of the **earth**, without **losing** your **dream** or your **imagination**. You **can** always **choose** to **switch** on or off your **PC** and **go** on **living** and **dreaming** in the *traditional* **way**. Obviously , when you **use** a **computer** and **play**, for **instance**, with a virtual **game**, you **have** to **know** you **are** **using** the **imagination** of the **software** **programmer** who **designed** the **game**. But you **can***live* , at the same **time**, the **dream** of **being** into a strange **planet** where there **are** three **suns** and **moons**. The **science** **technology** and its **appliance** in the **industry** **relalized** many **dreams** of the **mankind**, such as the

MCI = 7.8

```
ed, has, brought, ∅, ∅, ∅, ∅, ing, ∅, _k/d_, ∅, ∅, ing, ∅, ed, _e/o_, _i/ou_,
_fe__t, were, ed, left, ∅, ed, was, t, s, ∅, ed, are, s, has, _ea/o__en, ∅,
∅, is, ing, is, ing, ing, is, was, ed, ∅, ∅, was, ∅, ∅, ∅, ∅, ∅, is, ∅, ∅, ∅,
∅, ∅, ∅, were, s, is, ing, ∅, ∅, are, ∅, ∅, ∅, ∅, ∅, ∅, ∅, ∅, _e/t, ing, is,
is, are, n, ∅, ∅, ∅, ∅, has, brought, ing, ∅, ed, ing
```

VERBS AND NOUNS IN TEXT

The economic **welfare** **reached** by most of the European **Countries** especially after the second **world** **war**, **has** **brought** in our **homes** all **kinds** of **comfort** which **have** **revolutioned** our **customs** and therefore our **mentality**. I **can** **start** by **mentioning** the **television**, the **washing** **machine**, the **dish** **washer**, the **telephone** and **lots** of more **things** which **have** **made** our **life** easy and comfortable and without them we **could** not **live** **today**. **Talking** to my **parents** who **have** **experienced** the **war** and that **got** married during the **sixties**, I **found** out that in those **years** **people** **felt** great **enthusiasm** towards **life** and **were** **optimists** for the **future**. **Reconstruction** and **industrialisation** **created** new **jobs** with the **consequence** that **lots** of **people** **left** the **country** to **move** to **town** which **offered** more **opportunities** especially for the young **generations**. It **was** a golden **period**, where **everybody** **dreamt** about a better future and about a better **society**. It **seems** impossible , but although we **have** now **reached** through **technology** a high **standard** of **life**, we **are** very **pessimists**. It **seems** as **progress** **has** **stolen** our **imagination** and therefore the **love** for small **things**. I **can** **give** few **examples** that such a **fact**: **television** **is** **becoming** lately the **killer** of **conversation** between **parents** and **children**; it **is** almost **disappearing** the **use** of **writing** nice **letters** to **friends**, since there **is** the **telephone**; when I **was** a **child** I **used** to **invent** **games** to **enjoy** with my **friends**. Our **imagination** **was** so vast that we **could** **play** wonderful **games** with simple **stones** or

Morphological complexity across languages and genres

Pallotti, G. (2018) La complessità morfologica: ricerca e didattica. *Incontri. Rivista europea di studi italiani.* 33(1), 9–26.

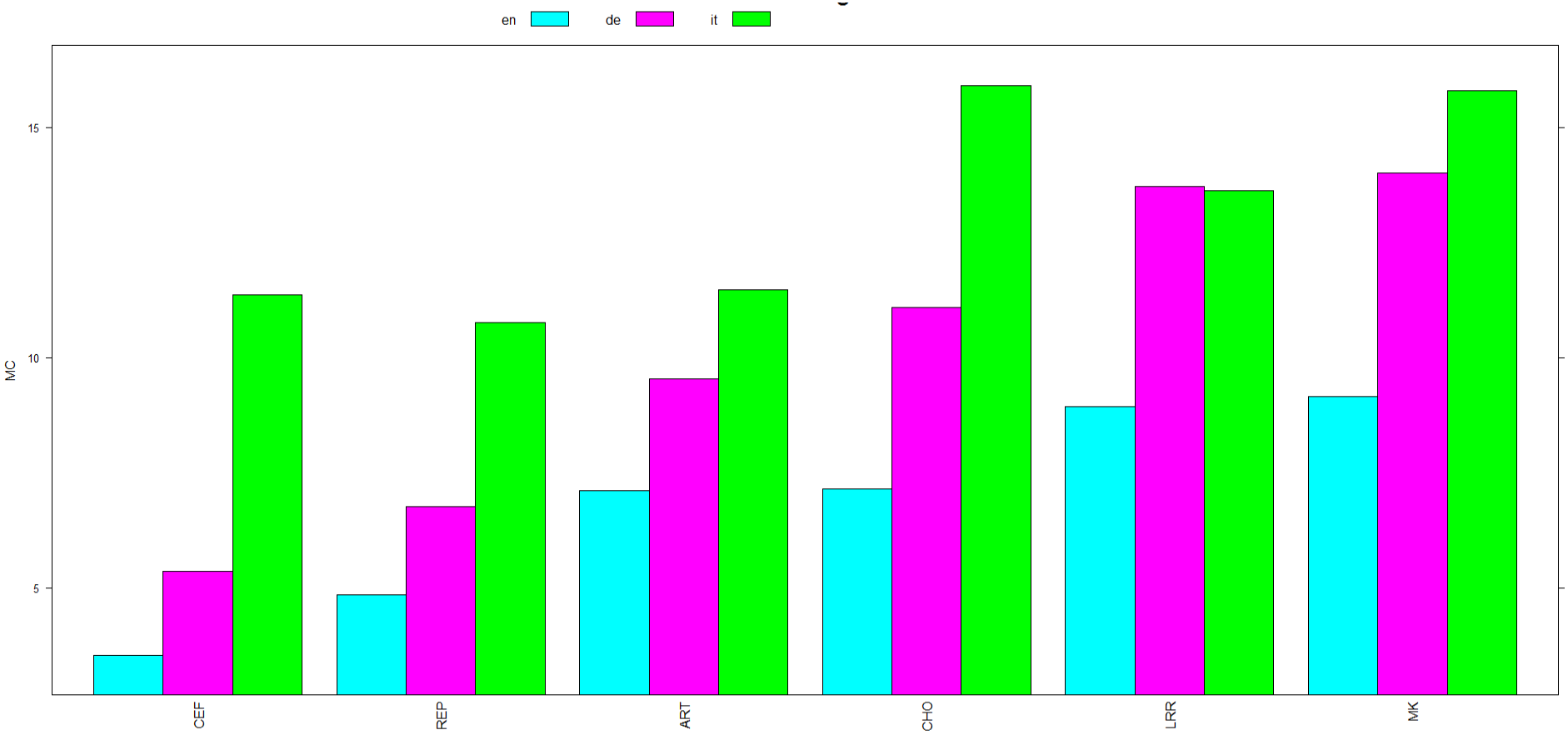
Data

1. Common European Framework of Reference for languages, general descriptor scales (CEF)
2. Little Red Riding Hood (LRRH)
3. N. Chomsky ‘What Uncle Sam really wants’ (sec 1-2) (CHO)
4. Mark’s gospel, ch 1-2 (MK)
5. EU press release (ART)
6. EU report on energy (items 1-10) (REP)

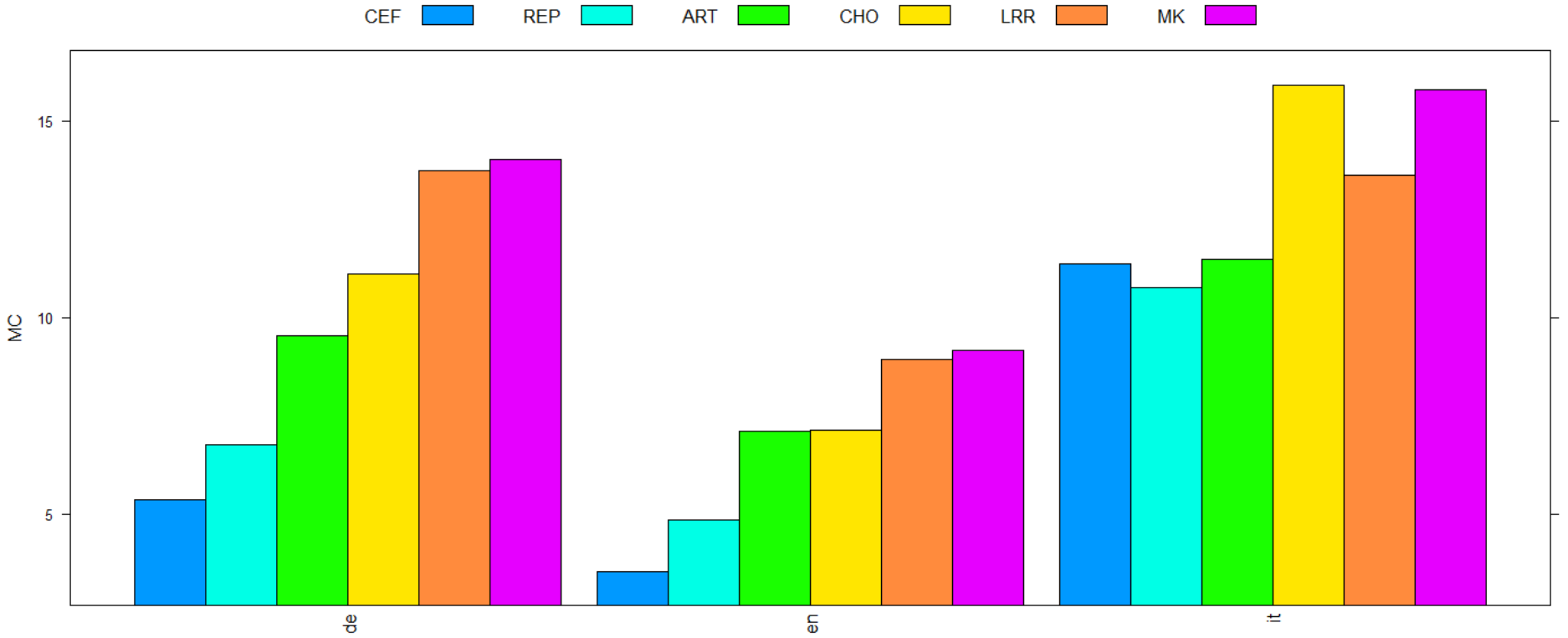
Parallel versions in three languages: English, German, Italian

All texts about 1,000 words

MC across texts



MC across languages



Range
German: 5.37-14.02 (=8.65)
English: 3.54-9.16 (=5.62)
Italian: 10.76-15.91 (=5.15)

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Two studies on MC in written L2 texts

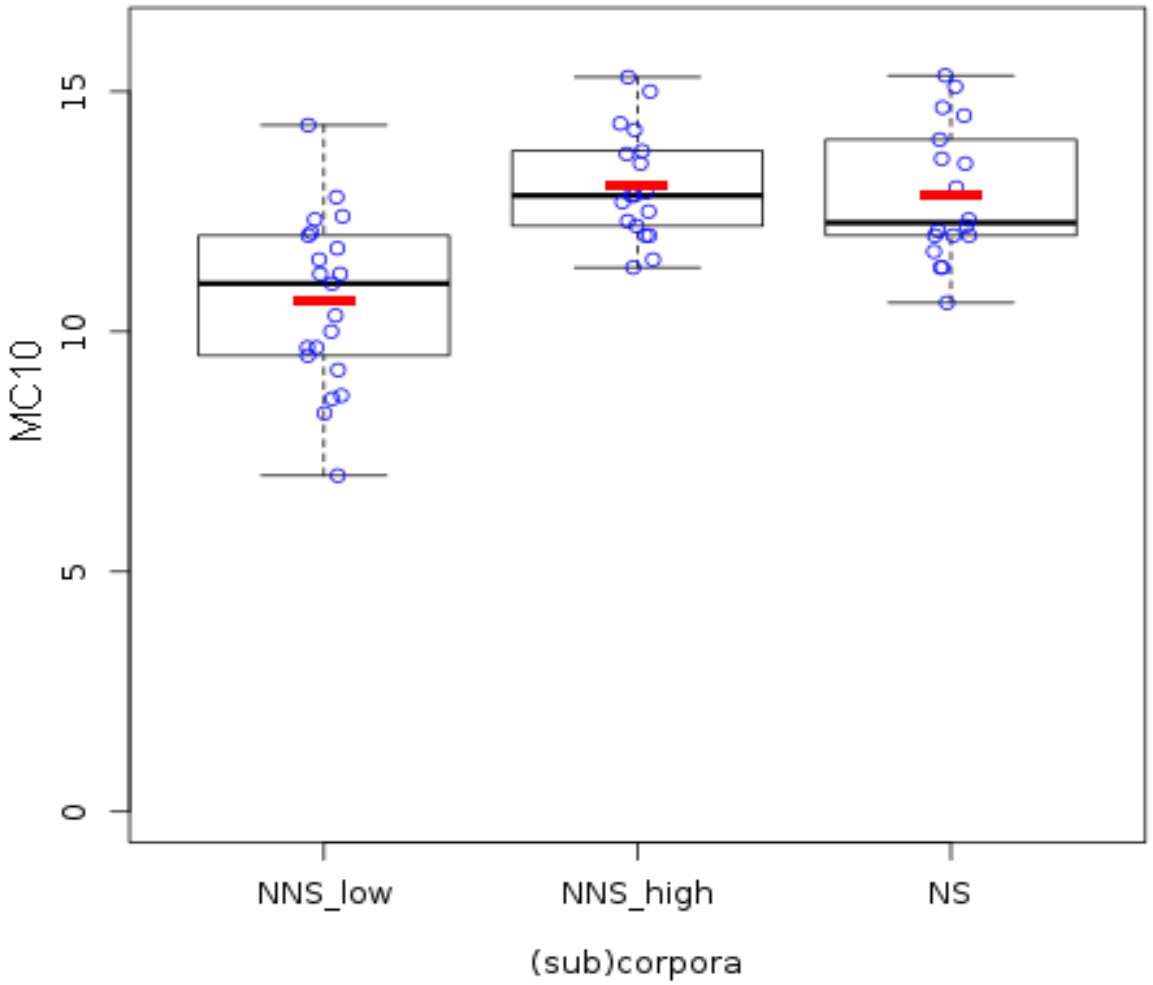
Brezina, V and Pallotti, G. (2019). Morphological complexity in written L2 texts. *Second Language Research*, 35, 99-119.

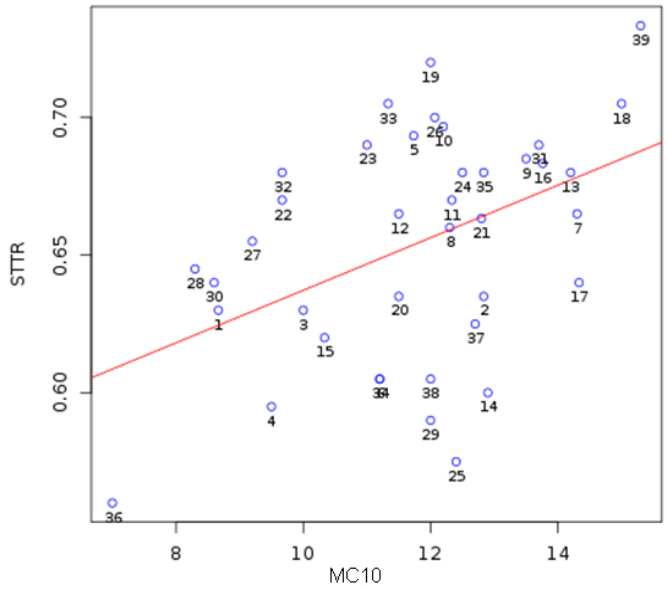
Study 1. MC in L2 Italian

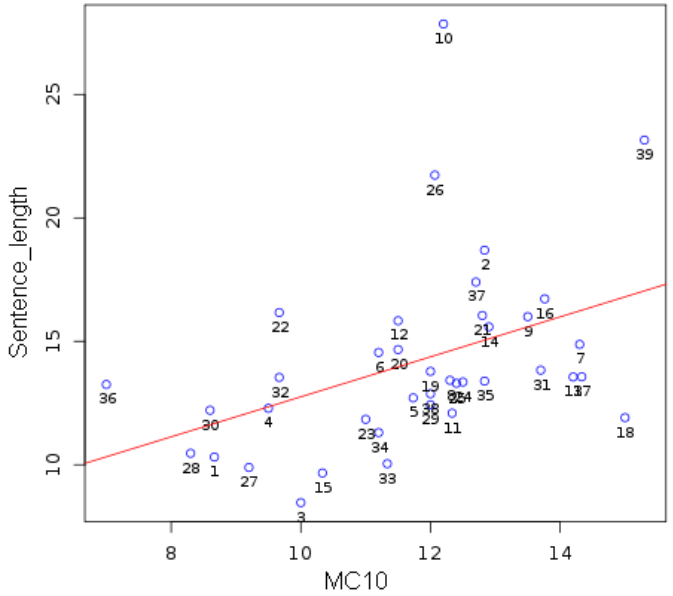
Two written argumentative essays per participant
NNS: Dutch university students learning Italian as a foreign language (A2-B2)
NS: native-speaking Italian university students
Measures: MCI, C-test, lexical complexity (STTR on 100 words), syntactic complexity (sentence length).

Corpus from project: 'Communicative Adequacy and Linguistic Complexity in L2 Writing' (CALC) (Kuiken, Vedder & Gilabert, 2010; Kuiken & Vedder, 2014).

Corpus	No. of texts	Corpus size (words)	Mean text length (*SD*)	Mean verbs/text (*SD*)
NNS	39	9,793	251 (54)	51.00 (10.94)
NS	18	4,384	244 (63)	42.78 (11.49)







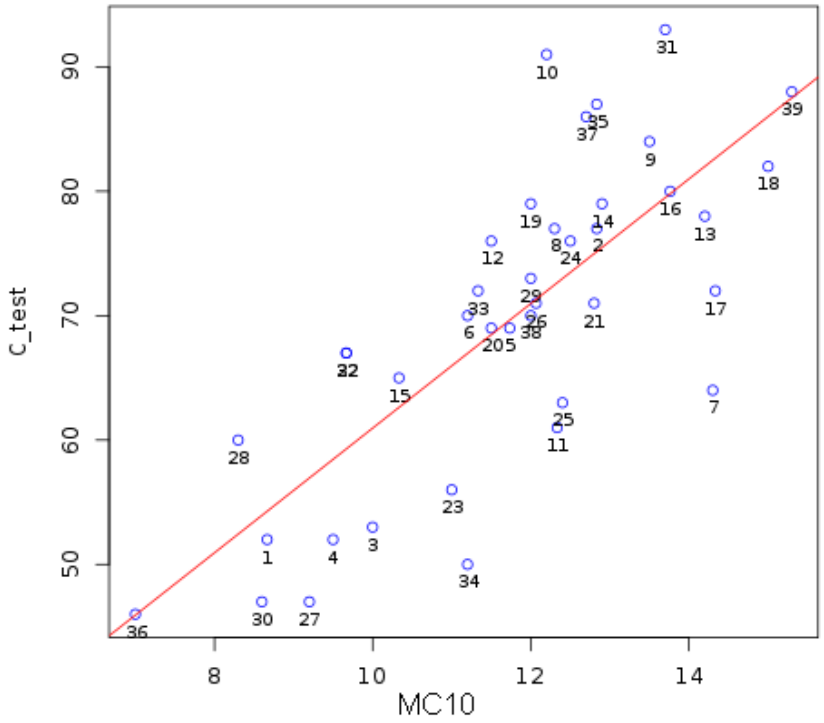
NNS data

MC10 correlates with:

- lexical complexity (STTR): r = 0.441, p = 0.005, 95% CI [0.145, 0.664]), medium effect size;

- syntactic complexity (sentence length): r = 0.416, p = 0.008, 95% CI [0.115, 0.646], medium effect size

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA



NNS data

Correlation between MC10 and proficiency (C-test): r = 0.759, p < 0.001, 95% CI [0.584, 0.867]

Study 2. MC in L2 English

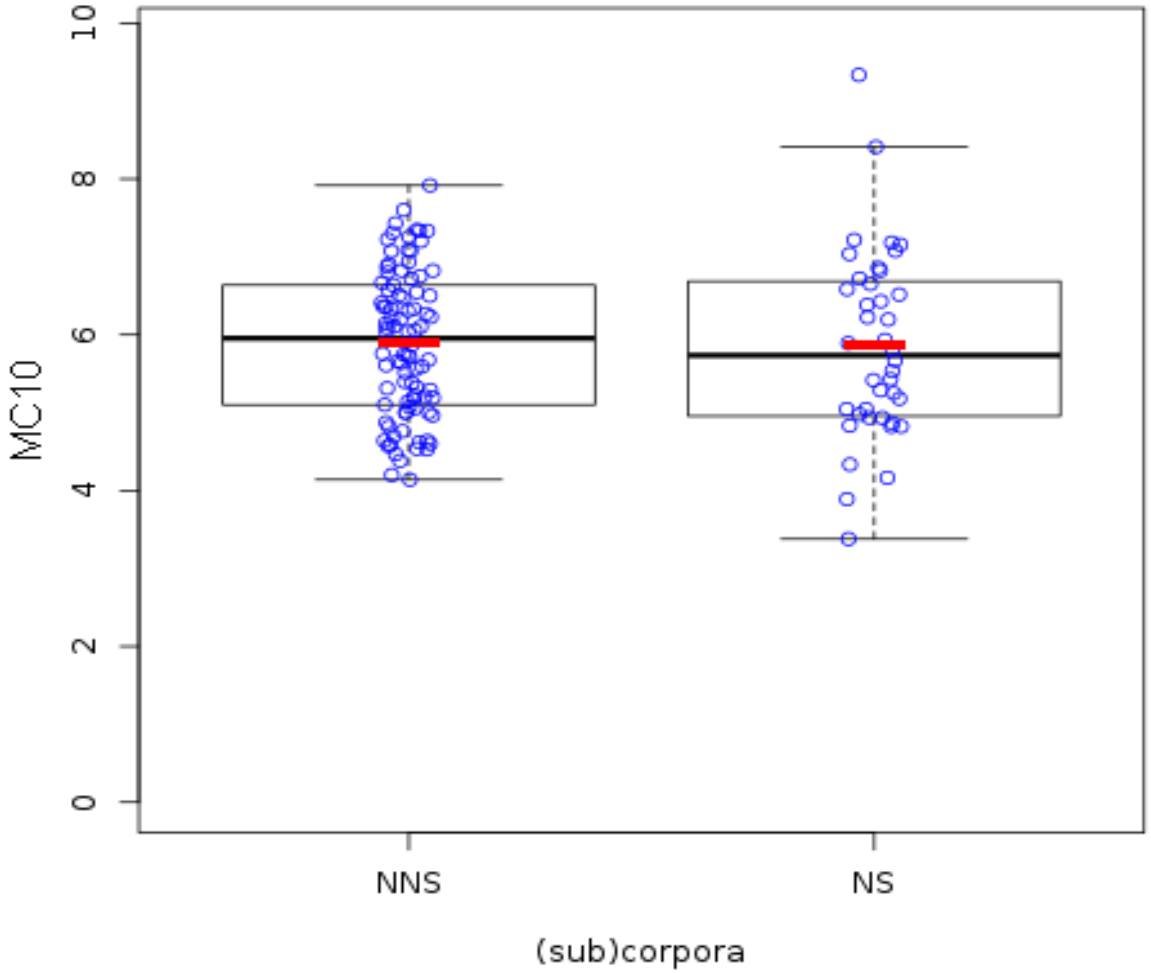
Argumentative essays in English

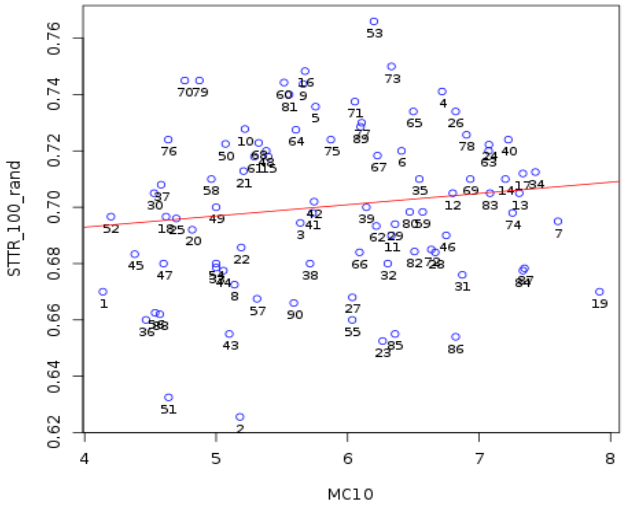
NNS: Italian university students (B1- C1)
NS: British and American university students.

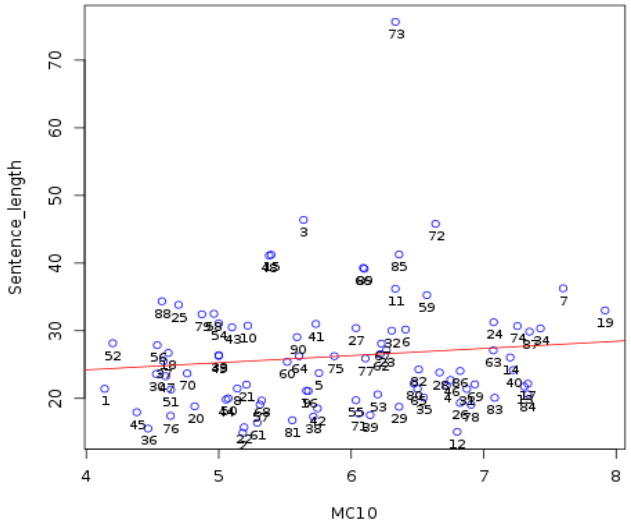
Measures: MCI, lexical complexity (STTR on 100 words), syntactic complexity (sentence length).

Corpora: ICLE (Granger et al., 2002), Locness (Granger. n. d.)

Corpus	No. of texts	Corpus size (words)	Mean text length (*SD*)	Mean verbs/text (*SD*)
NS	40	21,718	543 (103)	110.43 (30.29)
NNS	90	53,068	590 (191)	112.32 (32.48)







MCI does not statistically correlate with:

Lexical complexity (STTR): r = 0.158, p = 0.138, 95% CI [-0.051, 0.353];

Syntactic complexity (sentence length): r = 0.112, p = 0.295, 95% CI [-0.098, 0.312]

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Complexity: the more, the better?

Many studies report that lexical, morphological and syntactic complexity grow over time and are associated to higher scores in holistic ratings (e.g. Bulté & Housen, 2018; Crossley et al., 2011; Lahuerta Martínez, 2018; Yang et al., 2015)

Teaching how to write 'complex language' is a common goal for language teachers.

So it would seem that, in general, the more (complexity), the better.

But, is this always the case?

Studies questioning the ‘the more, the better’ assumption

Not ‘more complexity’ but ‘different complexity’

Biber et al (2011, 2016, 2020 etc.): oral conversation gets more complex with more subordination, written academic language complexity is mainly due to higher phrasal embedding.

Not ‘more complexity’ but ‘different complexity’

In syntax....

Ortega (2003), Norris & Ortega (2009): 1) cx increases as coordination; 2) cx increases as subordination; 3) cx increases in phrases

Nippold et al. (2005): in L1A, subordination rate remains constant between age 11 and 29, but nominal clauses (*I think...*) decrease while relative clauses increase.

Lambert & Nakamura (2018): L2 learners: beginners use more nominal clauses, advanced learners more adverbial and relative clauses.

Not ‘more complexity’ but ‘different complexity’

In text cohesion...

Crossley et al (2011): younger L1 writers (and those receiving lower scores) tend to use more cohesive devices, while more advanced writers produce more complex phrases.

In the lexicon...

Durrant & Brenchley (2019, 2022): younger L1 children use more low-frequency nouns (e.g., *caldron, fairy, hideout, wisp*), but tend to repeat them more often.

Less complexity may be better

“It is frequently the case that expert speakers and writers express complex ideas more simply than novices. This is not due to the availability of linguistic resources but rather to practiced mastery in efficient and effective message formation” (Lambert & Kormos 2014: 612)

Less complexity may be better

A study on syntactic complexity and discourse appropriateness in L1/L2 Italian
(Pallotti, in preparation)

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

VIP corpus

VIP (Variabilità nell'Interlingua Parlata; 'Variability in spoken interlanguage')

Participants

- 12 NNS female high school students, 4-8 yrs in Italy
- 10 NS female high school students

Data collection

- Longitudinal, 4 yrs; similar (but not identical) tasks every year (NS recorded only once or twice);
- Each participant performs a variety of communicative tasks: interview, map task, film retelling, planning a school trip, gathering information about electronic devices by making **phone calls**

Valentina (NS)

AG2: moito viaggi #
VAL: buonasera
AG2: sì
VAL: volevo chiederle informazione #
AG2: sì
VAL: e: per londra #
AG2: sì
VAL: una classe di: ragazzi
AG2: mh mh
VAL: e: qualcosa di conveniente che:: # che c'è

AG2: moito viaggi #
VAL: good evening
AG2: yes
VAL: I'd like to ask you information #
AG2: yes
VAL: er: to london #
AG2: yes
VAL: a class of youths
AG2: mh mh
VAL: er: something unexpensive that:: # that's available

UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Shirley (NNS)

AG3: moito viaggi?
SHI:.hhh #0_8 buonasera
#0_5
AG3: buonasera
SHI:allora noi siamo un # gruppo di quattro # amici che # vorremmo far un viaggio # a londra
#0_4
AG3: [sì
SHI: [perciò volevamo chiedere il costo dell'aerio e gli # orari #0_5

AG3: moito viaggi?
SHI:.hhh #0_8 good evening
#0_5
AG3: good evening
SHI:now we're a # group of four # friends who # would like to go on a trip # to london
#0_4
AG3: [yes
SHI: [so we wanted to ask the price of the plane and the # schedule #0_5

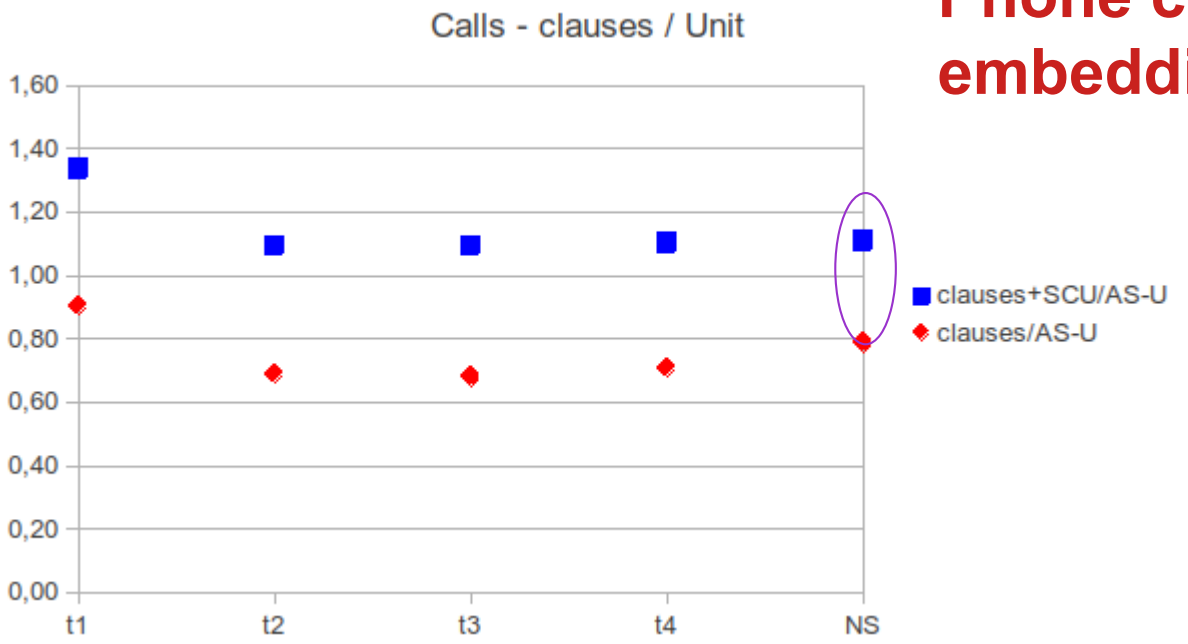
UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

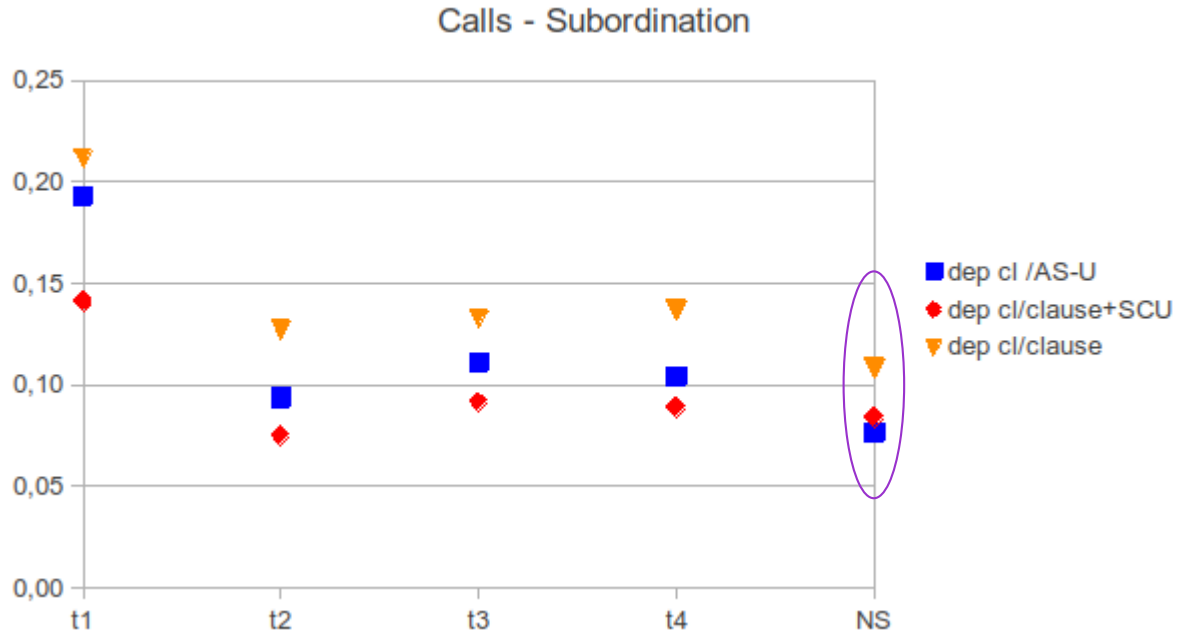
Aisha (NNS)

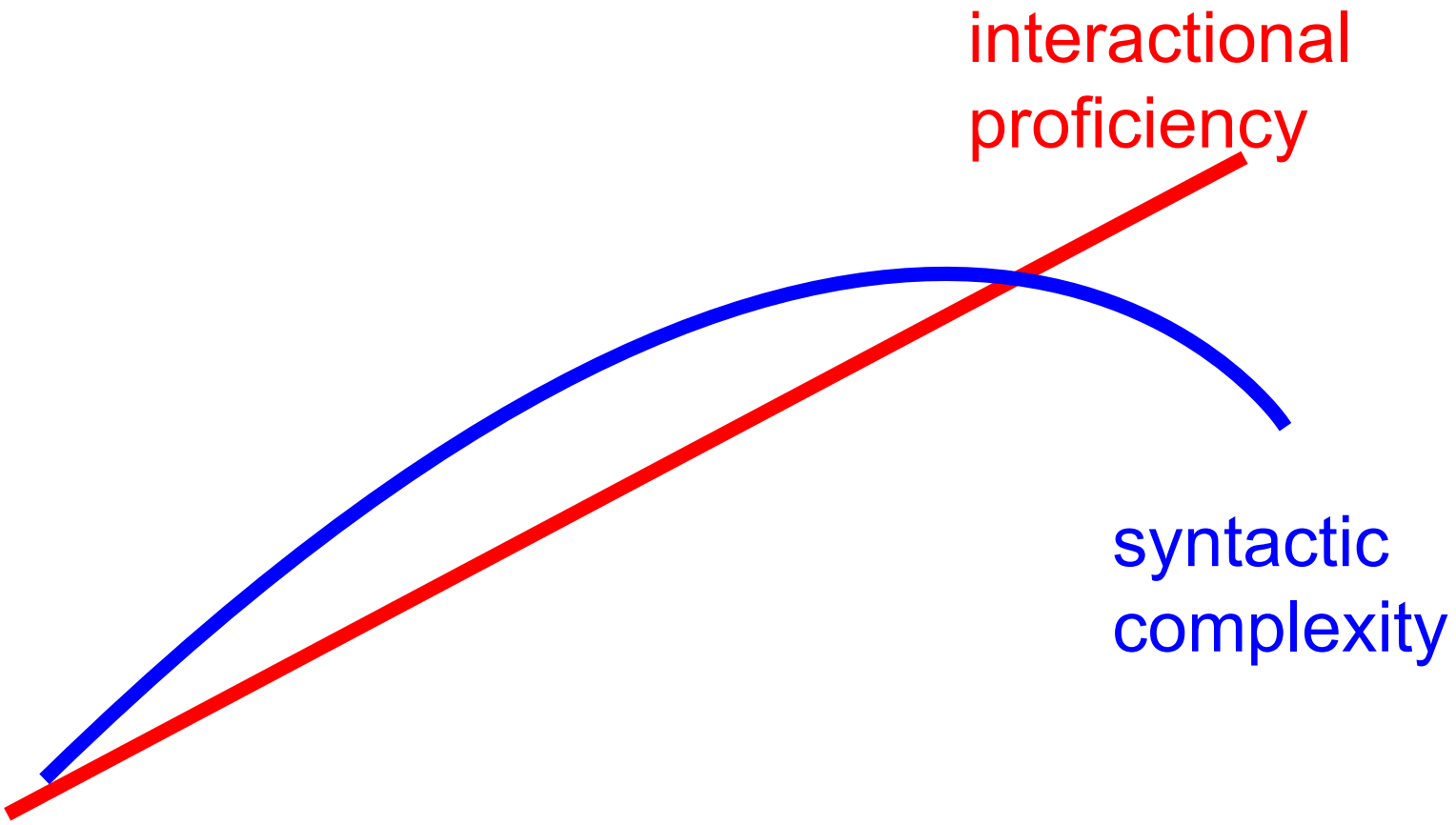
AG1: ci bi esse buongiorno sono Daniela.
ST4: buongiorno. ho bisogno dell'informazione per andare a barcellona

AG1: ci bi esse good morning Daniela speaking.
ST4: good morning. I need the information to go to barcelona

Phone calls – syntactic embedding







"'More complex' does not necessarily mean 'better'. Progress in a learner's language ability for use may include syntactic complexification, but it also entails the development of discourse and sociolinguisic repertoires that the language user can adapt appropriately to particular communication demands". (Ortega 2003: 494)

"The main conclusion seems to be that linguistic complexity grows when this is specifically required by the task and its goals, and not for the sake of it, as if learners aimed at complexification by default". (Pallotti 2009: 596)

Functional adequacy of linguistic complexity

Appropriate complexity

Nowadays, measuring complexity is not a problem (at least for English and a few other languages)

The real problem is interpreting complexity

Complexity needs to be interpreted, among other things, in terms of functional adequacy (Pallotti 2009)

“there are numerous grammatical devices associated with complexity, and so texts can be complex in very different ways in addition to being complex to differing extents” (Biber et al 2016: 648)

What is appropriate complexity in language production?

Register / genre / task appropriateness: what is appropriate for an academic paper may not be appropriate for a telephone call

Language-specific appropriateness: what is appropriate for English may not be so for German or Italian

What is appropriate complexity in language production?

The complexity levels found in top language performers

- studies using native speaker controls (e.g. Biber et al 2016, 2020; Michel et al 2019)

The complexity levels associated with high quality ratings

- studies correlating complexity values with proficiency ratings (e.g. Bi & Jiang 2020; Lahuerta Martínez 2018)

Pedagogical implications

Going beyond ‘the more, the better’

Academic essays should not be the ‘gold standard’ of language education.

Even academic essays may not always benefit from more complexity.

Textual connectives, subordinate clauses, nominalized phrases must be taught and learned, but should not become an obsession or a goal in themselves.

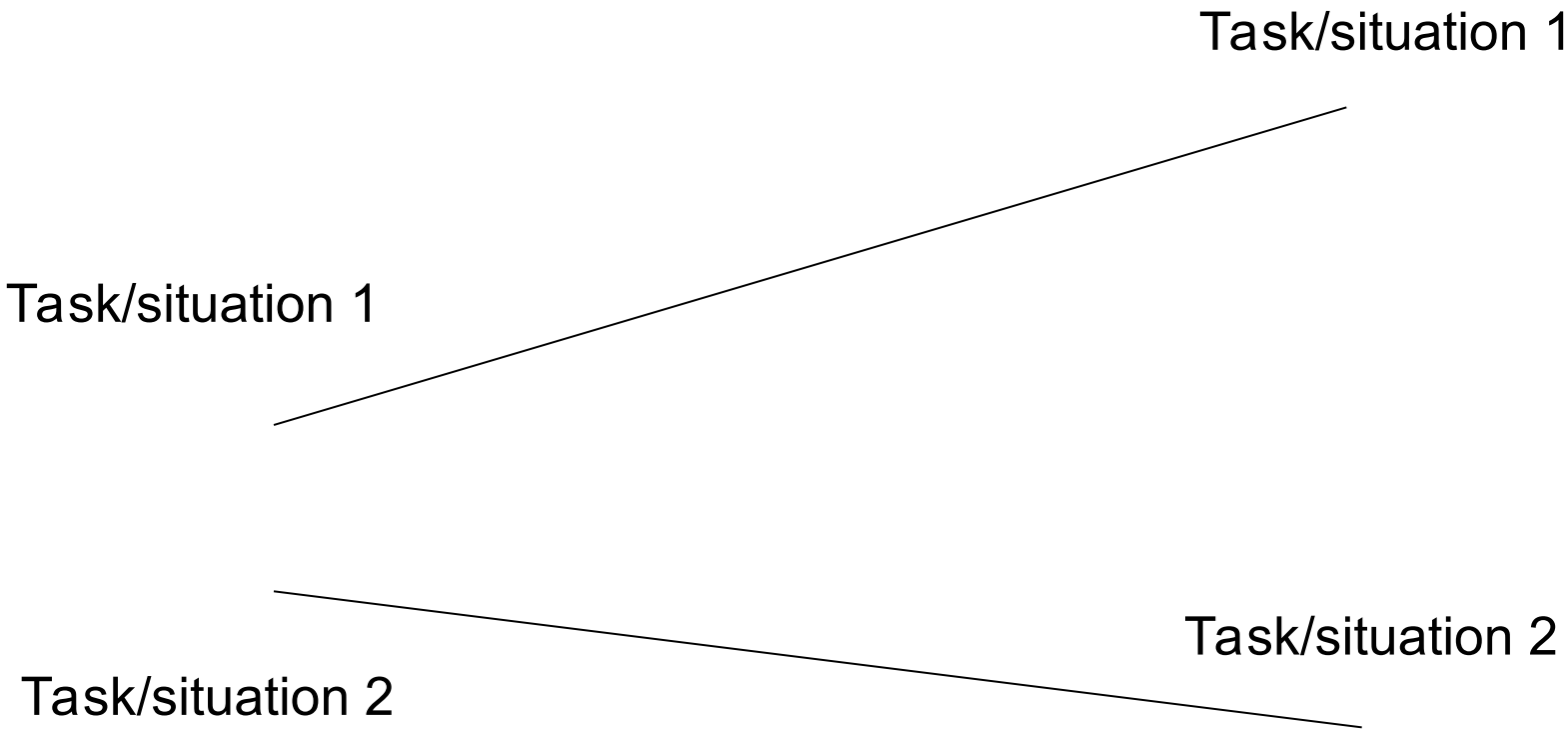
Rather than ‘teaching complexity’ we should teach appropriateness (to task, register, situation...).

Adolescent speakers of Swedish

	Verbal complexity in interviews	Difference between oral interviews and written texts
NNS	3.19	0.67
NS	2.40	1.00

(Wiklund, 2002)

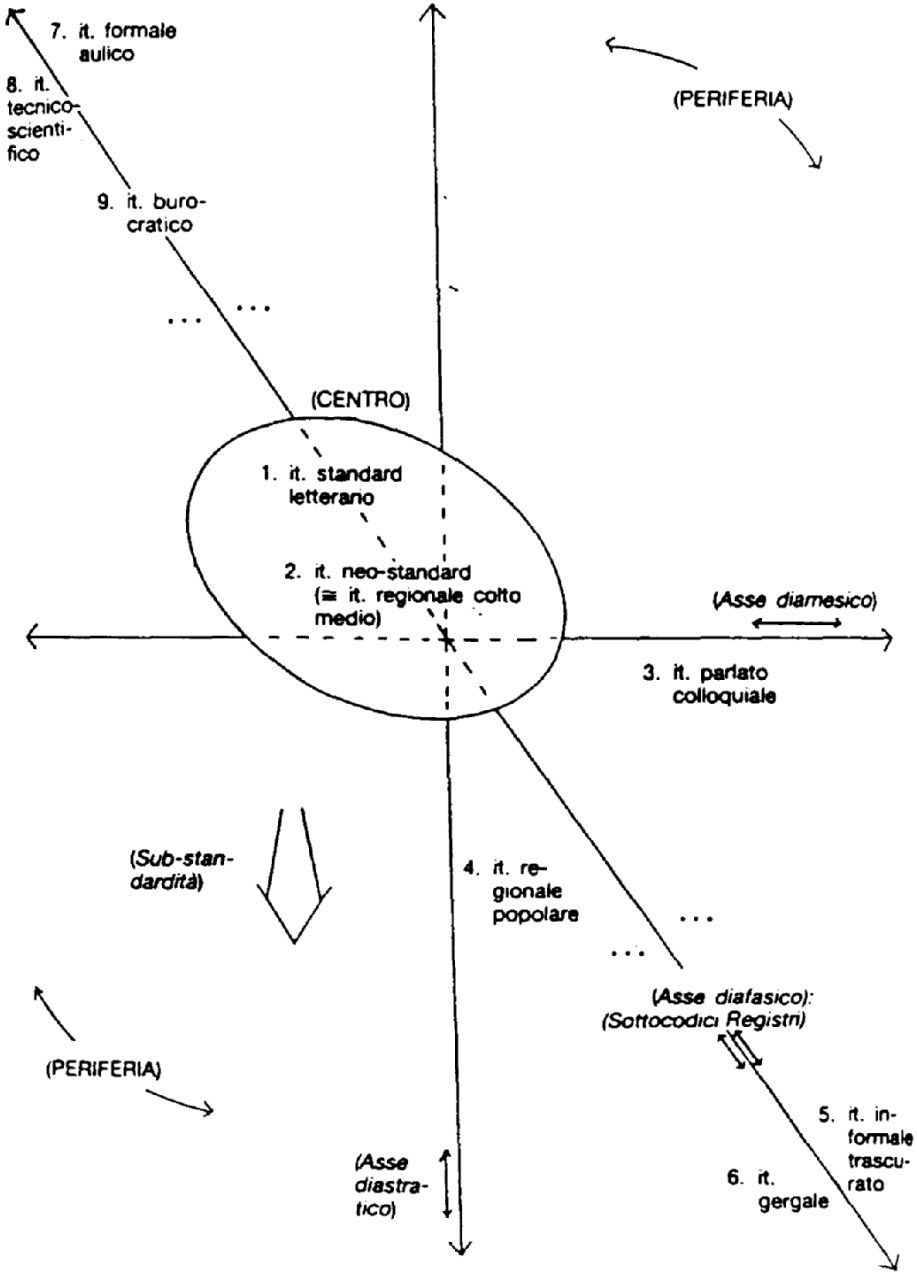
Teaching and learning variation



Beginner learners

Top language performers

Broadening the ‘language space’

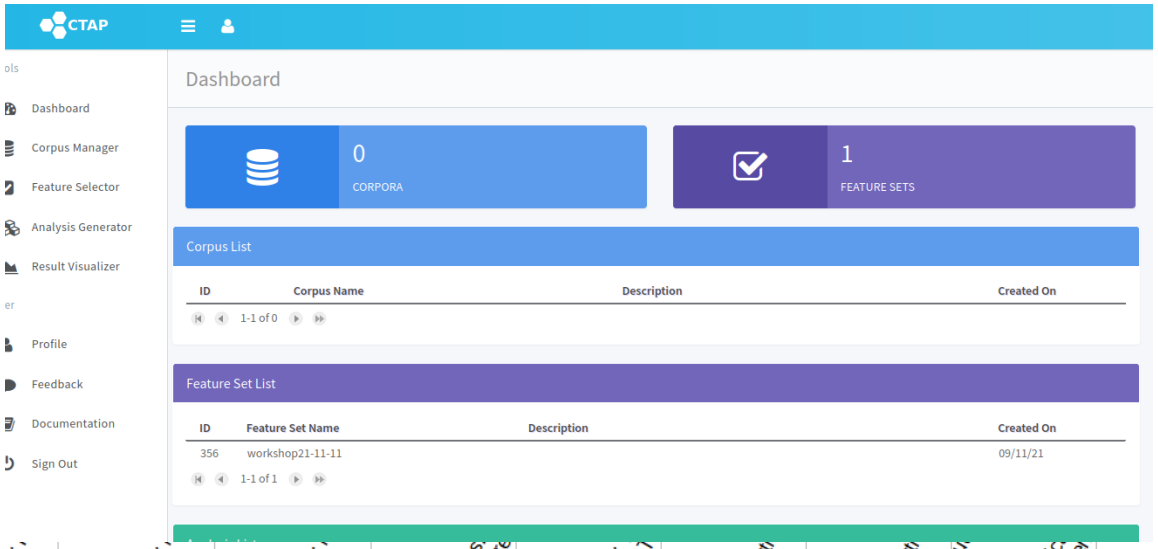


Berruto 1987

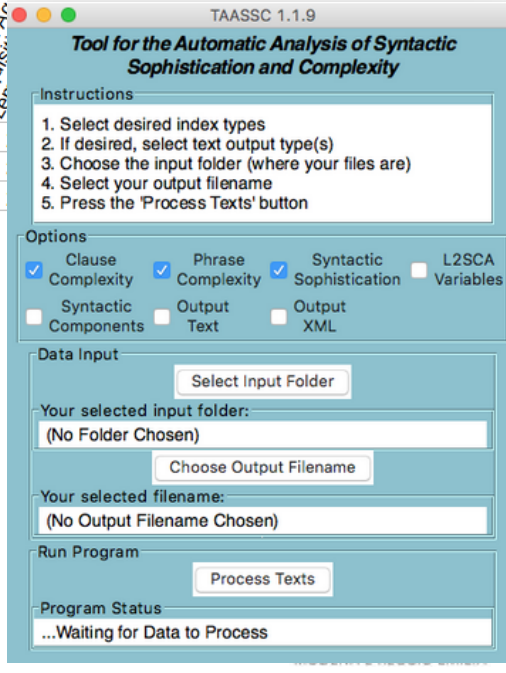
Teaching complexity awareness

Syntactic structures
- Word count (W)
- Sentence (S)
- Verb phrase (VP)
- Clause (C)
- T-unit (T)
- Dependent clause (DC)
- Complex T-unit (CT)
- Coordinate phrase (CP)
- Complex nominal (CN)

Syntactic complexity indices
- Mean length of sentence (MLS)
- Mean length of T-unit (MLT)
- Mean length of clause (MLC)
- Clause per sentence (C/S)
- Verb phrase per T-unit (VP/T)
- Clause per T-unit (C/T)
- Dependent clause per clause (DC/C)
- Dependent clause per T-unit (DC/T)
- T-unit per sentence (T/S)
- Complex T-unit ratio (CT/T)
- Coordinate phrase per T-unit (CP/T)
- Coordinate phrase per clause (CP/C)
- Complex nominal per T-unit (CN/T)
- Complex nominal per clause (CN/C)



Text_id	Tags	Text_	Lexical Richness: Lemma	Lexical Richness: Token	Lexical Richness: Token Ratio (Corrected TTR)	Lexical Richness: Token Ratio (Root TTR)	Lexical Sophistication Feature: Age of Acquisition (All Lemmas)	Lexical Sophistication Feature: Age of Acquisition (Lemmas)
6571		RR01norm.txt	0,8	0,82	28,83	28,83	5,48	7,75	3,04
6572		RR02norm.txt	0,79	0,83	31,3	34,78	6,04	8,54	3,27
6573		RR03norm.txt	0,79	0,81	33,27	33,27	5,59	7,91	2,62



Thank you!

the floor is yours...

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Grammatical vs Stylistic complexity

Grammatical complexity: complexity of grammatical rules

the rules for constructing a subordinate clause in German are more complex than those for constructing one in English, and they are obligatory

Stylistic complexity: complexity resulting from speaker's choices

the proportion of subordinate clauses in a German or English text is a matter of speaker's choice, or of language-specific rhetorical preferences, but it is never mandatory (if a text contains many/few subclauses no rules are violated)

Study 2.
Complexity variation across tasks – a focus on top language performers

Research design

Data from VIP corpus, elicited with same methodology

Tasks investigated: interview, film retelling, map task with peer, phone calls, face-to-face negotiations during school trip organization

Participants: 10 adolescent female native speakers of Italian (Elisa, Valentina + 8 more)

Measures

Lexical complexity

- Moving-Average TTR (MATTR): mean TTR in 250-word sequential samples

- % of non-basic words (>2k)

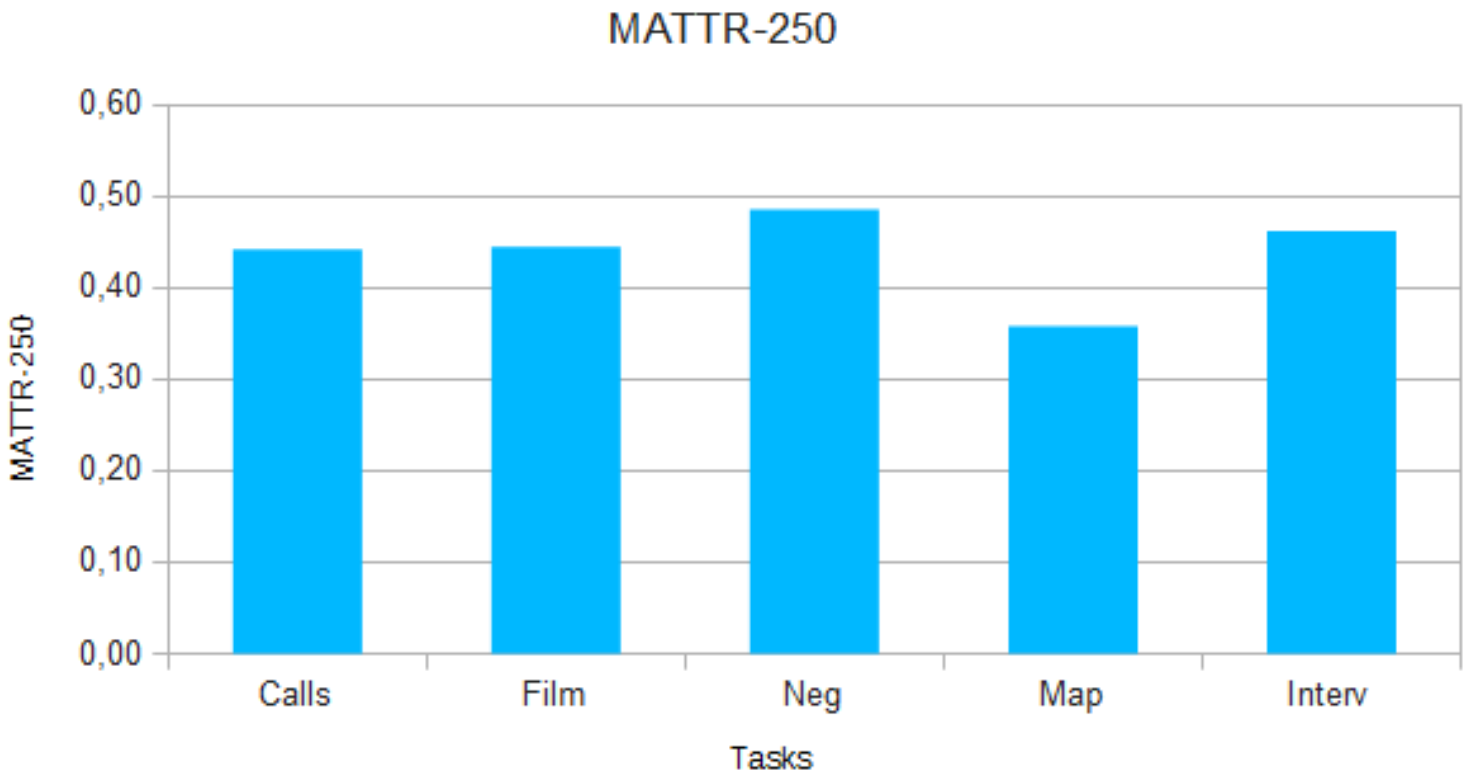
Morphological complexity

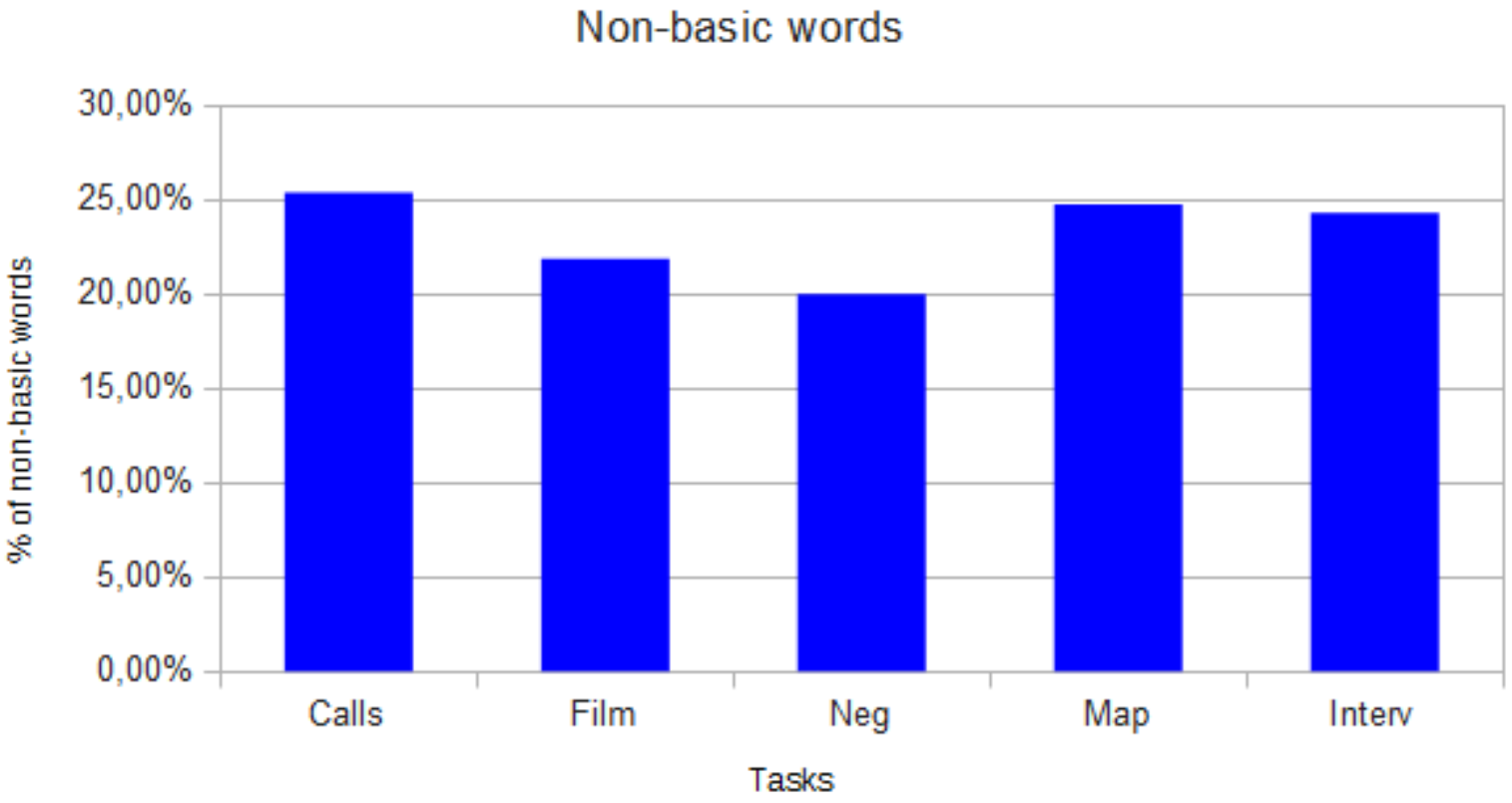
- Morphological Complexity Index, i.e. mean variety of verbal inflections within and across samples of 10 forms each, with 100 random sampling cycles

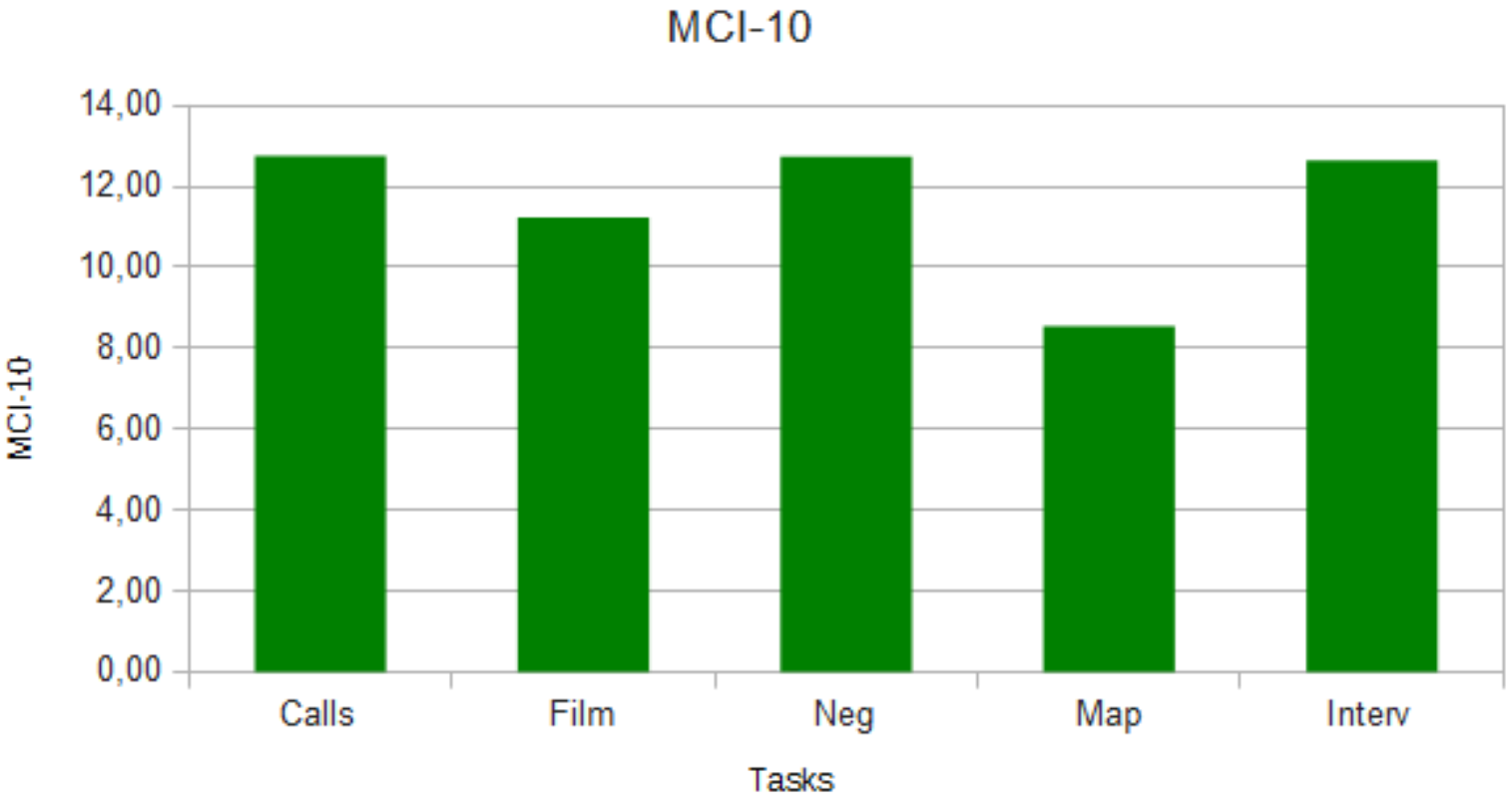
Syntactic complexity

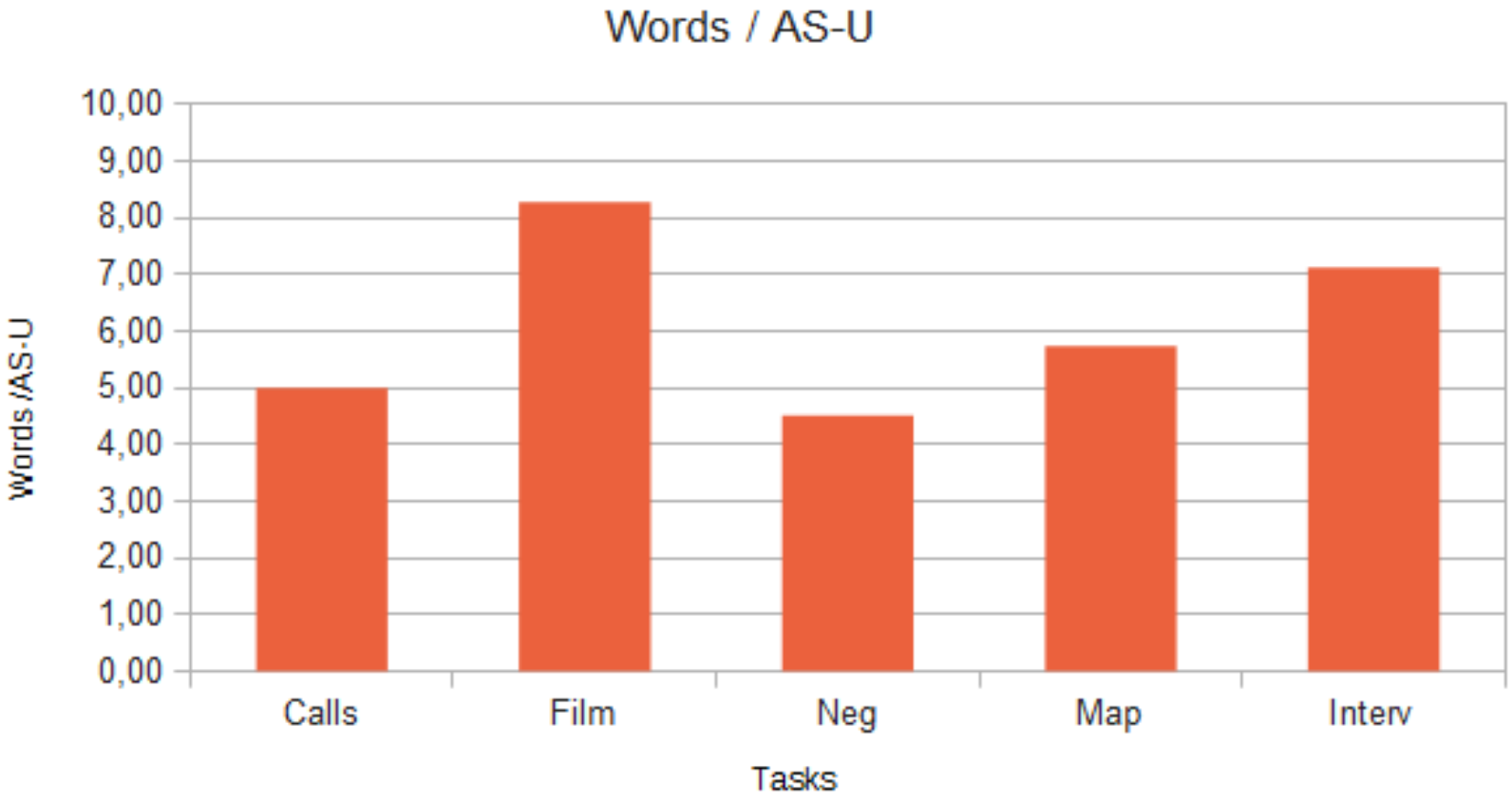
- Mean length of AS-Unit

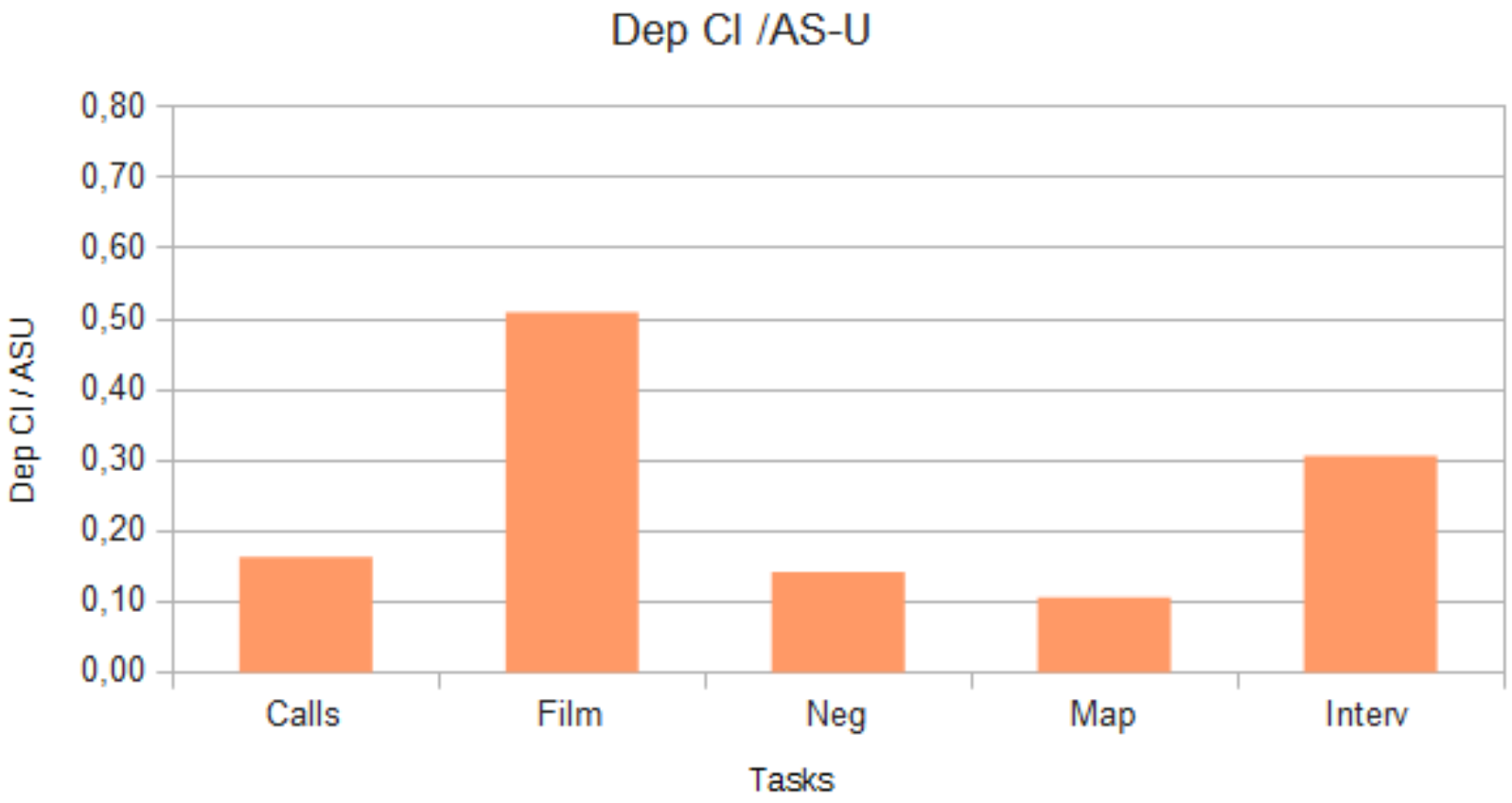
- Dependent clauses / AS-Unit

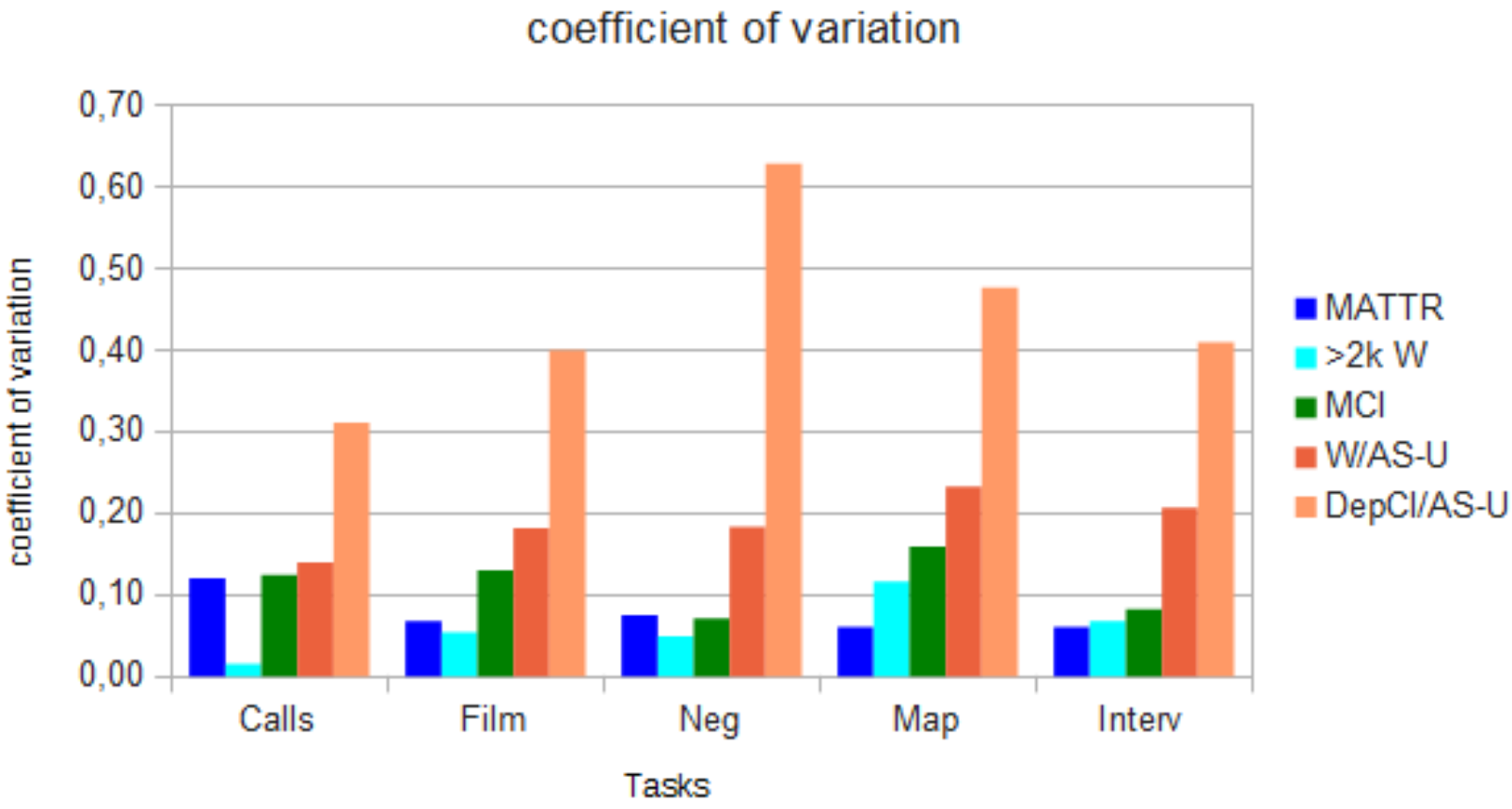












Results

Lexical complexity: small-medium variation across tasks, and relative independence of diversity (MATTR) from sophistication (rare words)

Morphological complexity: little variation across tasks, except for a rather low level in Map task

Syntactic complexity: clear variation across tasks, for both length of unit and clausal embedding

Inter-individual variation: small for lexical and morphological complexity, high for length of unit and very high for clausal embedding (individual style).