The syntax-morphology trade-off

Kilu von Prince
kilu.von.prince@hu-berlin.de

Language Science Kolloquium,
30.11.2017

Information density and typology

- Information density has been used as a proxy for linguistic complexity (Juola, 1998, 2008; Pellegrino *et al.* , 2011; Ehret & Szmrecsanyi, 2016; Koplenig *et al.* , 2017).

Information density and typology

- Information density has been used as a proxy for linguistic complexity (Juola, 1998, 2008; Pellegrino *et al.* , 2011; Ehret & Szmrecsanyi, 2016; Koplenig *et al.* , 2017).
- Questions concerning cross-linguistic differences in complexity at different levels are at the core of typological research and crucial for the relevance of typology beyond linguistics (Atkinson, 2011; Everett, 2005)

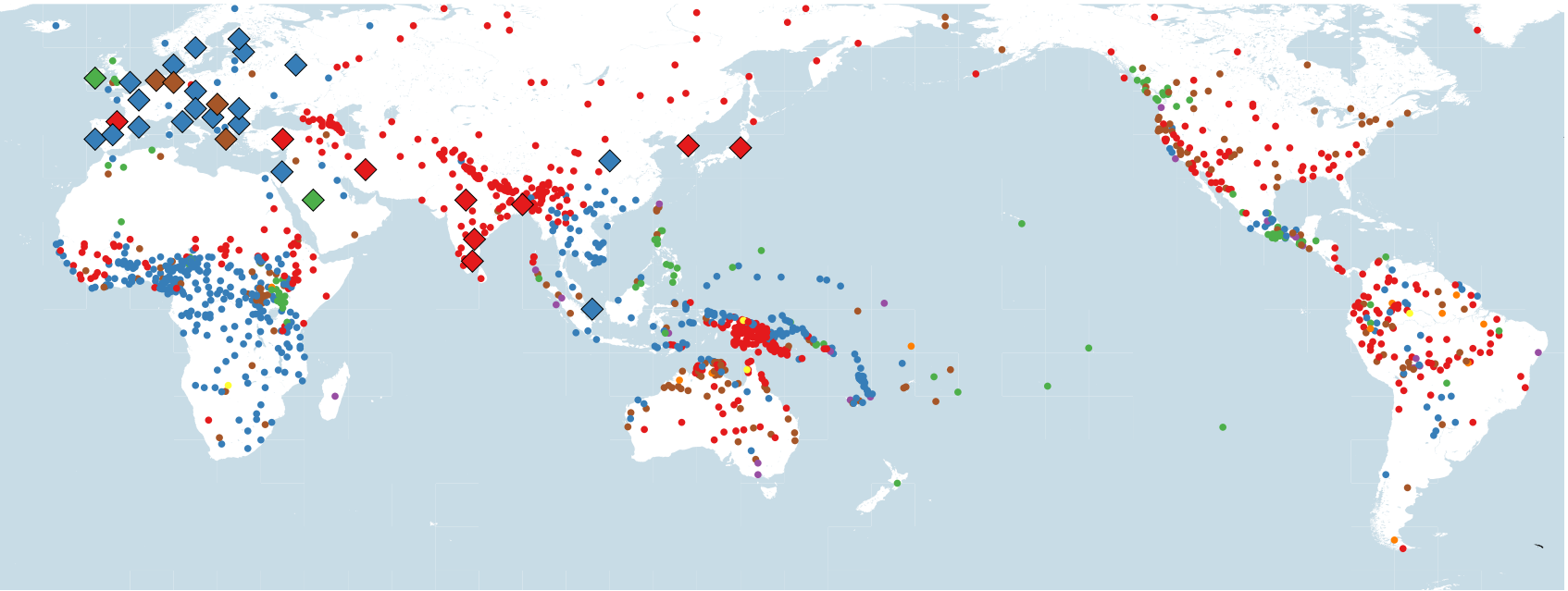
Information density and typology

- Information density has been used as a proxy for linguistic complexity (Juola, 1998, 2008; Pellegrino *et al.* , 2011; Ehret & Szmrecsanyi, 2016; Koplenig *et al.* , 2017).
- Questions concerning cross-linguistic differences in complexity at different levels are at the core of typological research and crucial for the relevance of typology beyond linguistics (Atkinson, 2011; Everett, 2005)
- Information density has also played a role in assessing properties of different word orders (Futrell *et al.* , 2015; Gildea & Jaeger, 2016).

Information density and typology

- Information density has been used as a proxy for linguistic complexity (Juola, 1998, 2008; Pellegrino *et al.* , 2011; Ehret & Szmrecsanyi, 2016; Koplenig *et al.* , 2017).
- Questions concerning cross-linguistic differences in complexity at different levels are at the core of typological research and crucial for the relevance of typology beyond linguistics (Atkinson, 2011; Everett, 2005)
- Information density has also played a role in assessing properties of different word orders (Futrell *et al.* , 2015; Gildea & Jaeger, 2016).
- There is still a certain gap between typological methodology and corpus-based investigations.

The challenge of inter(sub-)disciplinary research



Futrell *et al.* (2015) make a claim about universal word order properties, based on a sample of 37 languages, from 6 different families, all from Eurasia, with only 4 out of 7 possible word orders → not typologically representative (Rijkhoff & Bakker, 1998). Map data from Dryer (2013).

Basic observations

- Some languages express argument structure, information structure, sentence type and other categories primarily by morphological marking.

Basic observations

- Some languages express argument structure, information structure, sentence type and other categories primarily by morphological marking.
- Other languages use word structure for the same functions.

Basic observations

- Some languages express argument structure, information structure, sentence type and other categories primarily by morphological marking.
- Other languages use word structure for the same functions.
- Example: case marking vs. word order for expressing argument structure:

Basic observations

- Some languages express argument structure, information structure, sentence type and other categories primarily by morphological marking.
- Other languages use word structure for the same functions.
- Example: case marking vs. word order for expressing argument structure:

Basic observations

- Some languages express argument structure, information structure, sentence type and other categories primarily by morphological marking.
- Other languages use word structure for the same functions.
- Example: case marking vs. word order for expressing argument structure:

(1) a. den Hund hat die Katze gebissen

Basic observations

- Some languages express argument structure, information structure, sentence type and other categories primarily by morphological marking.
- Other languages use word structure for the same functions.
- Example: case marking vs. word order for expressing argument structure:

(1) a. den Hund hat die Katze gebissen
b. die Katze hat den Hund gebissen

Basic observations

- Some languages express argument structure, information structure, sentence type and other categories primarily by morphological marking.
- Other languages use word structure for the same functions.
- Example: case marking vs. word order for expressing argument structure:

(1) a. den Hund hat die Katze gebissen
b. die Katze hat den Hund gebissen

(2) the cat bit the dog

A brief history

- Nichols (1992) is the first to systematically quantify morphological complexity, in a sample of 200 languages, based on grammatical descriptions.

A brief history

- Nichols (1992) is the first to systematically quantify morphological complexity, in a sample of 200 languages, based on grammatical descriptions.
- Siewierska (1998) first statistically established the general trend for a syntax-morphology trade-off in argument marking, based on grammatical descriptions.

A brief history

- Nichols (1992) is the first to systematically quantify morphological complexity, in a sample of 200 languages, based on grammatical descriptions.
- Siewierska (1998) first statistically established the general trend for a syntax-morphology trade-off in argument marking, based on grammatical descriptions.
- The first comparative text-based assessment of morphological complexity is Juola (1998, 2008);

A brief history

- Nichols (1992) is the first to systematically quantify morphological complexity, in a sample of 200 languages, based on grammatical descriptions.
- Siewierska (1998) first statistically established the general trend for a syntax-morphology trade-off in argument marking, based on grammatical descriptions.
- The first comparative text-based assessment of morphological complexity is Juola (1998, 2008);
- Ehret & Szmrecsanyi (2016) took the methodology by Juola (1998), applied it to syntax as well and morphology in ten varieties of English and six other European languages.

A brief history

- Nichols (1992) is the first to systematically quantify morphological complexity, in a sample of 200 languages, based on grammatical descriptions.
- Siewierska (1998) first statistically established the general trend for a syntax-morphology trade-off in argument marking, based on grammatical descriptions.
- The first comparative text-based assessment of morphological complexity is Juola (1998, 2008);
- Ehret & Szmrecsanyi (2016) took the methodology by Juola (1998), applied it to syntax as well and morphology in ten varieties of English and six other European languages.
- Koplenig *et al.* (2017) used related methods to establish a broad cross-linguistic trend in close to 1200 languages from the Parallel Bible Corpus (Mayer & Cysouw, 2014).

Targeting structual levels: Koplenig *et al.* (2017)

Original	i called her yesterday and i called her today because i wanted to talk to her
Masked word structure	i itweiy khk doeerdsun rki i itweiy khk ehtuy ahuwlok i hwkilr dw weyy dw khk
Destroyed word order	her wanted i today talk i i her yesterday called because called her to to and

A closer look at methods: Measuring complexity

- Juola (1998) suggests that Zip-compression allows for an approximation of Kolmogorov complexity. Ehret & Szmrecsanyi (2016) follow.

A closer look at methods: Measuring complexity

- Juola (1998) suggests that Zip-compression allows for an approximation of Kolmogorov complexity. Ehret & Szmrecsanyi (2016) follow.
- Koplenig *et al.* (2017) take a different measure:

A closer look at methods: Measuring complexity

- Juola (1998) suggests that Zip-compression allows for an approximation of Kolmogorov complexity. Ehret & Szmrecsanyi (2016) follow.
- Koplenig *et al.* (2017) take a different measure:

Approximation to the entropy of a specific string

$H = \left[\frac{1}{N} \sum_{i=2}^{N} \frac{l_i}{log(i)} \right]^{-1}$, where l_i is 1 plus the length of the maximal string that can be predicted from the character at position i in that it has previously followed it.

A closer look at methods: Measuring complexity

- Juola (1998) suggests that Zip-compression allows for an approximation of Kolmogorov complexity. Ehret & Szmrecsanyi (2016) follow.
- Koplenig *et al.* (2017) take a different measure:

Approximation to the entropy of a specific string

$H = \left[\frac{1}{N} \sum_{i=2}^{N} \frac{l_i}{log(i)} \right]^{-1}$, where l_i is 1 plus the length of the maximal string that can be predicted from the character at position i in that it has previously followed it.

A closer look at methods: Measuring complexity

- Juola (1998) suggests that Zip-compression allows for an approximation of Kolmogorov complexity. Ehret & Szmrecsanyi (2016) follow.
- Koplenig *et al.* (2017) take a different measure:

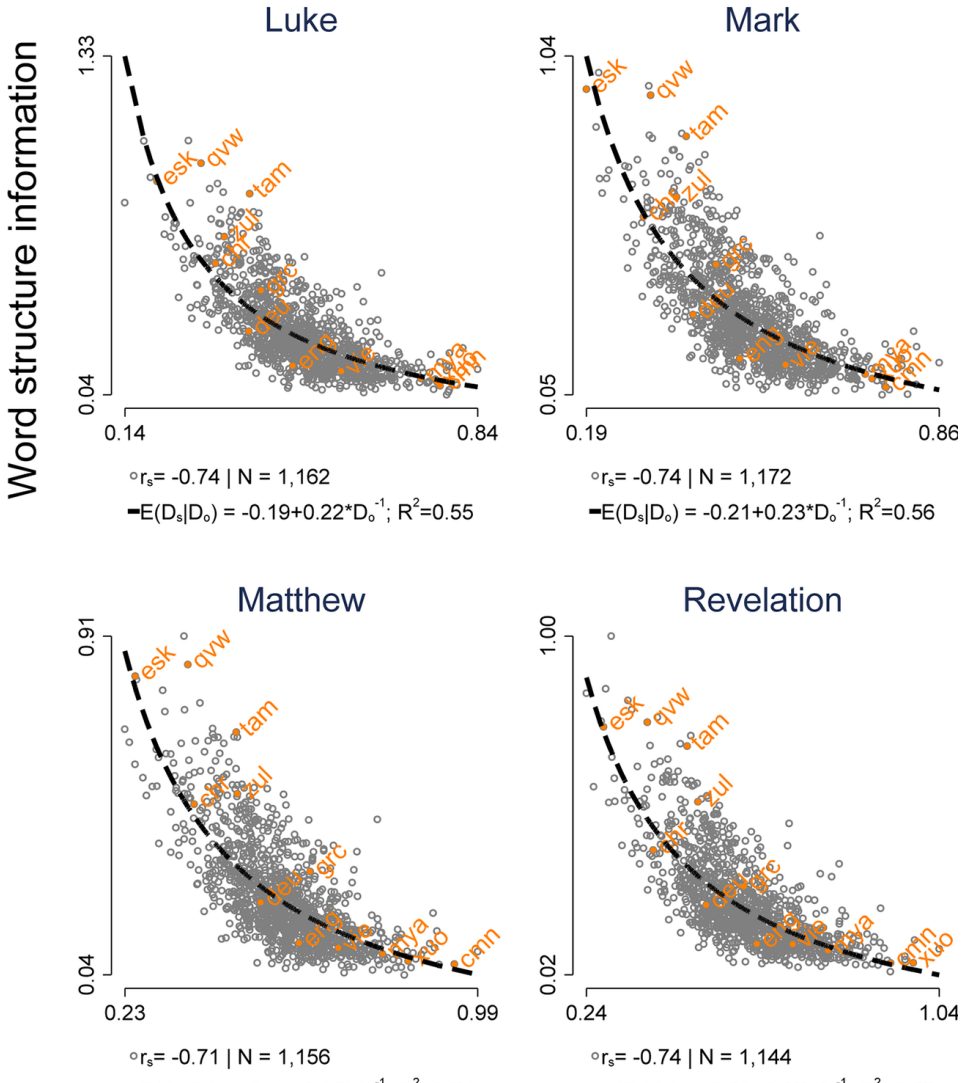
Approximation to the entropy of a specific string

$H = \left[\frac{1}{N}\sum_{i=2}^{N}\frac{l_i}{log(i)}\right]^{-1}$, where l_i is 1 plus the length of the maximal string that can be predicted from the character at position i in that it has previously followed it.

Example:

(3) a. they perceived that he spake of them
b. they supposed that they should have $\rightarrow l_i = 9$

Results


Word structure information
Luke
1.33
0.04
0.14
0.84
esk
qvw
tam
r_s= -0.74 | N = 1,162
$E(D_s|D_o) = -0.19+0.22*D_o^{-1}$; R^2=0.55
Mark
1.04
0.05
0.19
0.86
esk
qvw
tam
r_s= -0.74 | N = 1,172
$E(D_s|D_o) = -0.21+0.23*D_o^{-1}$; R^2=0.56
Matthew
0.91
0.04
0.23
0.99
esk
qvw
tam
r_s= -0.71 | N = 1,156
Revelation
1.00
0.02
0.24
1.04
esk
qvw
tam
r_s= -0.74 | N = 1,144

Effects of writing systems?

- By replacing a few thousand Chinese characters (types) with two dozen Latin letters, you may produce new regularities.

Effects of writing systems?

- By replacing a few thousand Chinese characters (types) with two dozen Latin letters, you may produce new regularities.
- We would expect that the writing system may have a significant effect on the relevant values of a language.

Effects of writing systems?

- By replacing a few thousand Chinese characters (types) with two dozen Latin letters, you may produce new regularities.
- We would expect that the writing system may have a significant effect on the relevant values of a language.

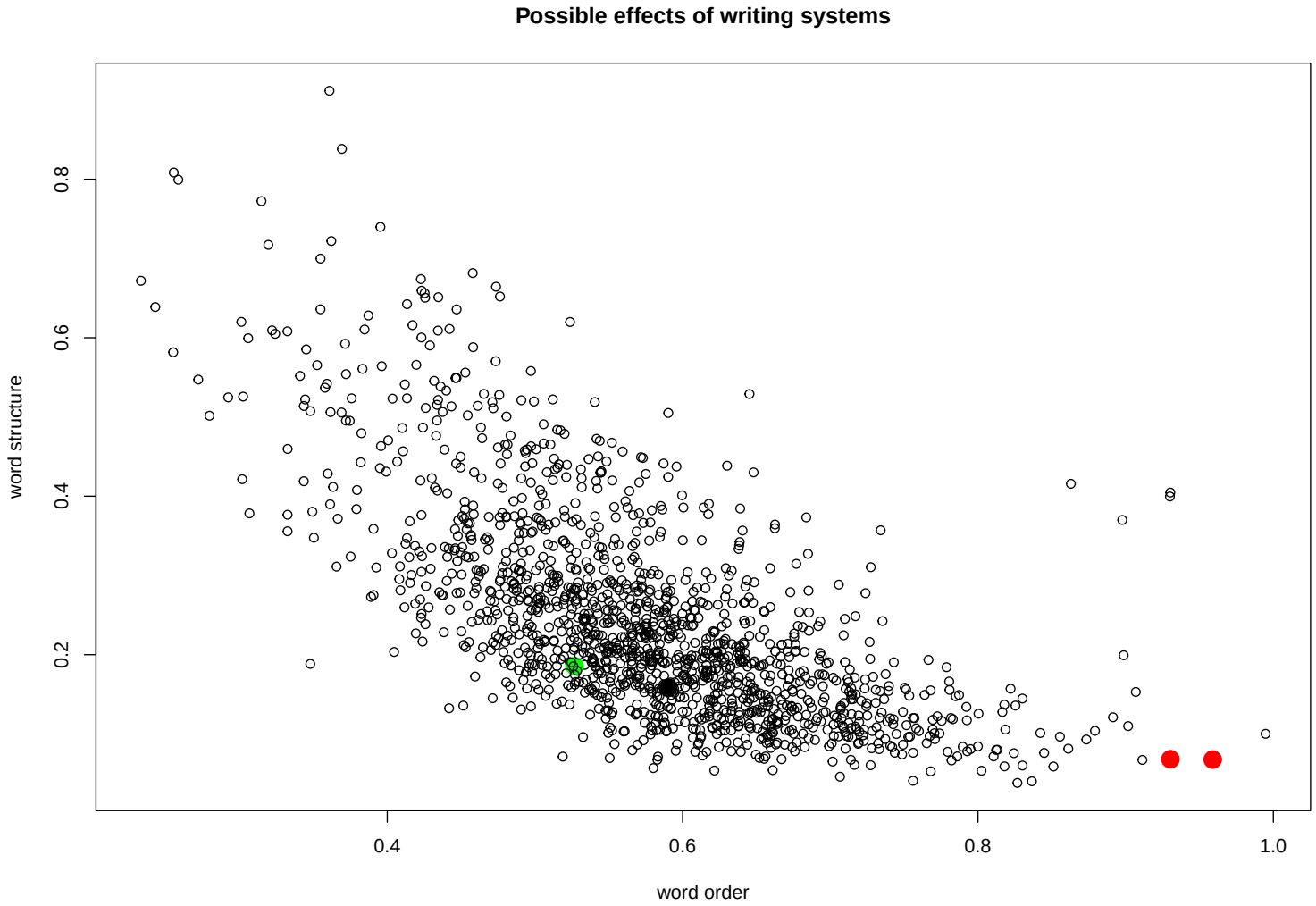
Effects of writing systems?

- By replacing a few thousand Chinese characters (types) with two dozen Latin letters, you may produce new regularities.
- We would expect that the writing system may have a significant effect on the relevant values of a language.

(4) Hakka Chinese
Yâ-sû Kî-tuk he Thai-ví ke heu-thoi , Thai-ví he Â-pak-lâ-hón ke heu-thoi

(5) Mandarin Chinese
耶穌 基督 —— 大衛 的 後裔 、 亞伯拉罕 的 後代 —— 他 的 家譜

Illustration of possible writing system effects



Non-linear morphology

- Not all morphological processes are strictly linear;
- Non-linear morphological processes are also known as fusional;
- They may involve
  - complex stem alternations (inflection);
  - consonantal, sceletal roots;
  - suprasegmental processes (tone);
  - vowel harmony;

Illustration of non-linear morphological processes

Consonantal roots: Hebrew (Semitic)

g d r "lock in"

gadar "he locked in"

gudar "he was locked in"

Tonal morphology: Kisi (Atlantic)

(6) a. *Ò cìmbù.*
3SG leave.PRES.HABITUAL
"She (usually) leaves."

b. *Ò cìmbú.*
3SG leave.PST.PFV
She left.

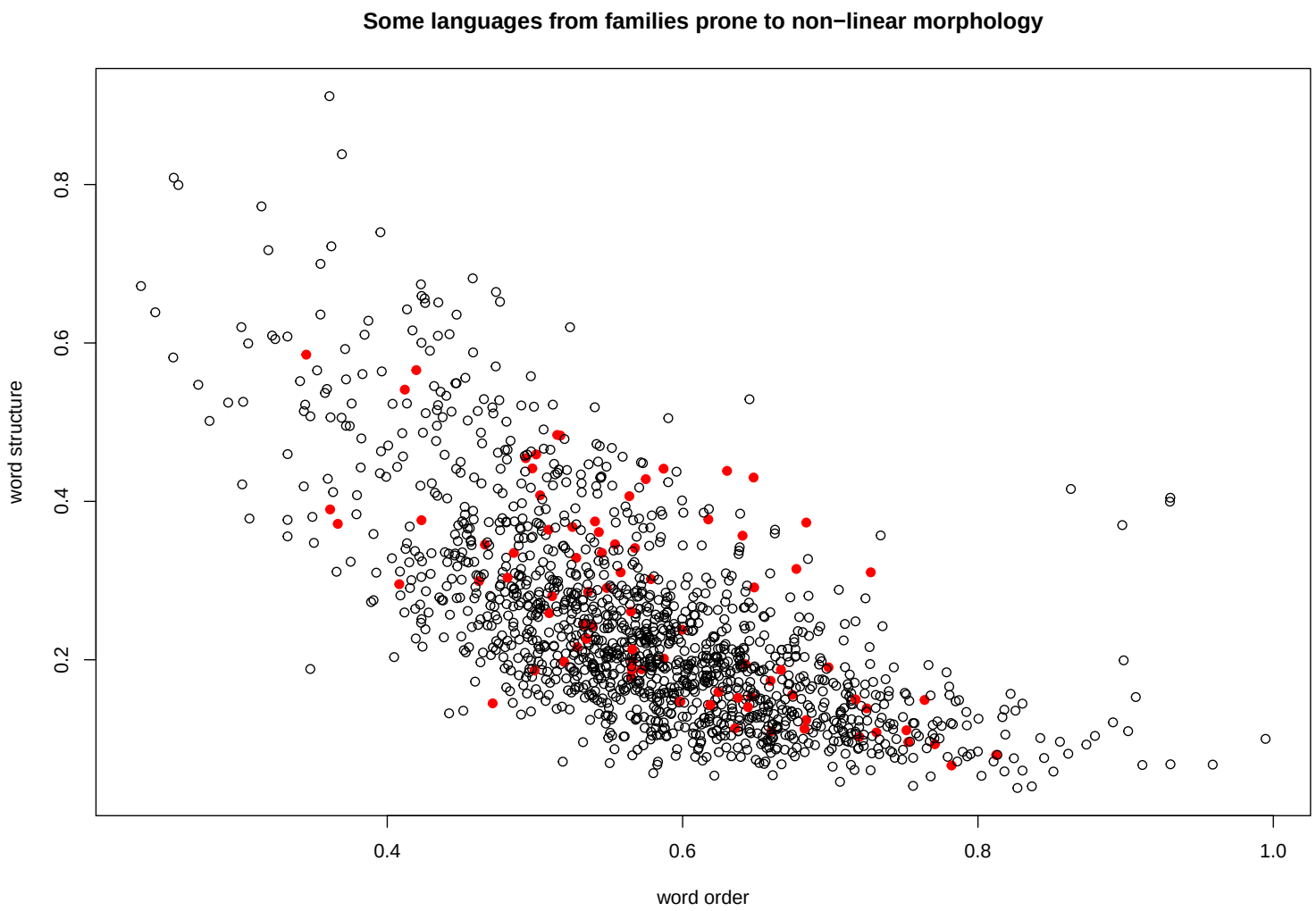
Non-linear morphology in Koplenig *et al.* (2017)

- The method of measuring I_i to determine approximate entropy levels relies on unbroken character sequences.

→ non-linear processes cannot be taken into account.

→ one expectation is that languages with non-linear morphology are outliers.

Possible effects of non-linear morphology



Further questions

- What other limitations does the process proposed by Koplenig *et al.* (2017) have?

Further questions

- What other limitations does the process proposed by Koplenig *et al.* (2017) have?
  - How does phonological complexity (inventory size, suprasegmentals, phonotactics) affect the results of this process?

Further questions

- What other limitations does the process proposed by Koplenig *et al.* (2017) have?
  - How does phonological complexity (inventory size, suprasegmentals, phonotactics) affect the results of this process?
  - What happens if two languages with similar morphological properties encode different amounts of information grammatically?

Further questions

- What other limitations does the process proposed by Koplenig *et al.* (2017) have?
  - How does phonological complexity (inventory size, suprasegmentals, phonotactics) affect the results of this process?
  - What happens if two languages with similar morphological properties encode different amounts of information grammatically?
  - How does lexical inventory size affect the process?

Further questions

- What other limitations does the process proposed by Koplenig *et al.* (2017) have?
  - How does phonological complexity (inventory size, suprasegmentals, phonotactics) affect the results of this process?
  - What happens if two languages with similar morphological properties encode different amounts of information grammatically?
  - How does lexical inventory size affect the process?
- Can we find ways to quantify these different structural levels separately?

Further questions

- What other limitations does the process proposed by Koplenig *et al.* (2017) have?
  - How does phonological complexity (inventory size, suprasegmentals, phonotactics) affect the results of this process?
  - What happens if two languages with similar morphological properties encode different amounts of information grammatically?
  - How does lexical inventory size affect the process?
- Can we find ways to quantify these different structural levels separately?
- To the extent that the process is reliable, which hypotheses can be tested with it?

Further questions

- What other limitations does the process proposed by Koplenig *et al.* (2017) have?
  - How does phonological complexity (inventory size, suprasegmentals, phonotactics) affect the results of this process?
  - What happens if two languages with similar morphological properties encode different amounts of information grammatically?
  - How does lexical inventory size affect the process?
- Can we find ways to quantify these different structural levels separately?
- To the extent that the process is reliable, which hypotheses can be tested with it?
  - Correlations with creole vs. non-creole languages?

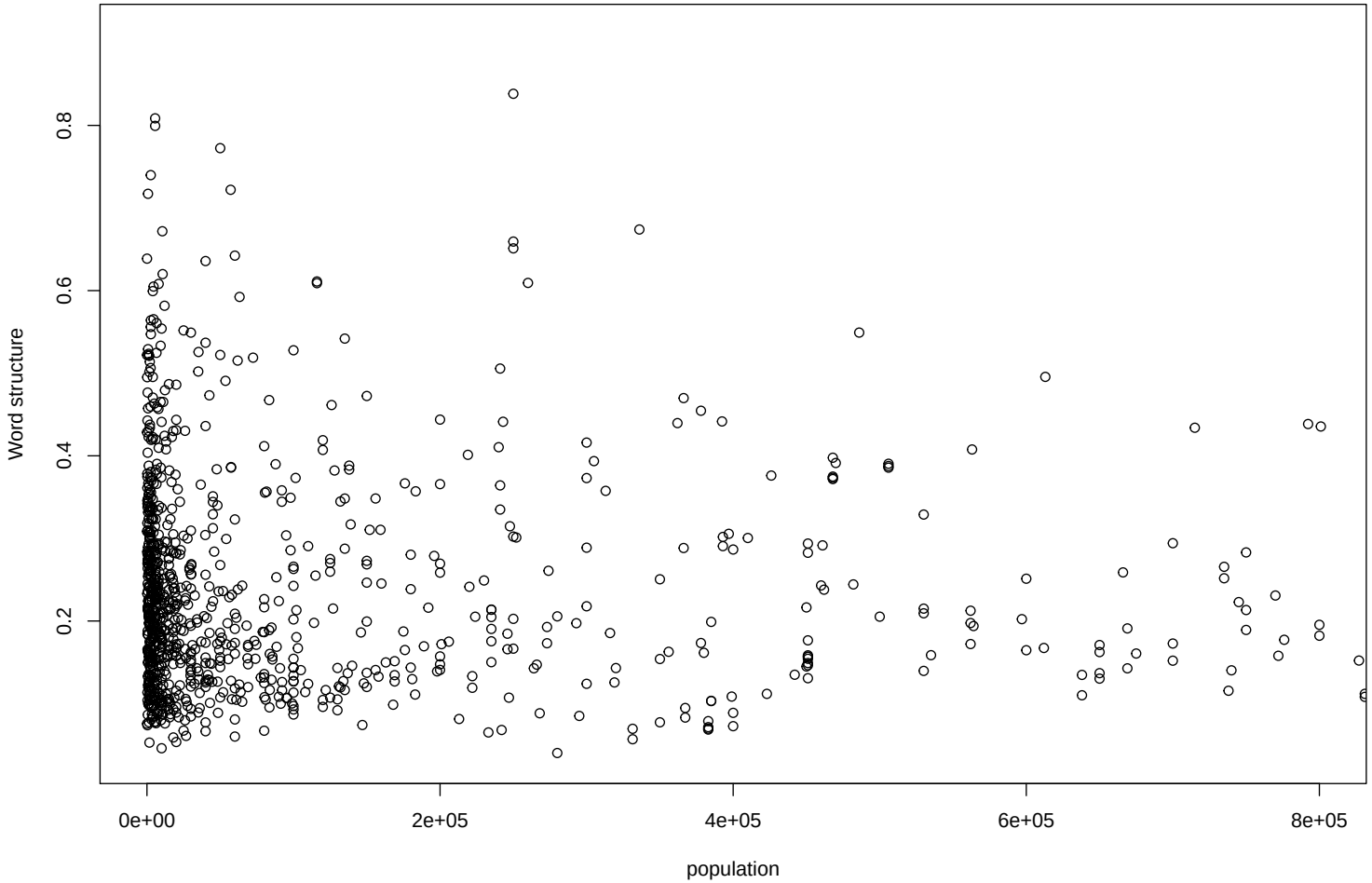
Further questions

- What other limitations does the process proposed by Koplenig *et al.* (2017) have?
  - How does phonological complexity (inventory size, suprasegmentals, phonotactics) affect the results of this process?
  - What happens if two languages with similar morphological properties encode different amounts of information grammatically?
  - How does lexical inventory size affect the process?
- Can we find ways to quantify these different structural levels separately?
- To the extent that the process is reliable, which hypotheses can be tested with it?
  - Correlations with creole vs. non-creole languages?
  - Correlations with population size/ number of L2 learners?

Further questions

- What other limitations does the process proposed by Koplenig *et al.* (2017) have?
  - How does phonological complexity (inventory size, suprasegmentals, phonotactics) affect the results of this process?
  - What happens if two languages with similar morphological properties encode different amounts of information grammatically?
  - How does lexical inventory size affect the process?
- Can we find ways to quantify these different structural levels separately?
- To the extent that the process is reliable, which hypotheses can be tested with it?
  - Correlations with creole vs. non-creole languages?
  - Correlations with population size/ number of L2 learners?
  - Correlations with geographic features?

Population size and word structure



Yet another perspective

- There are probably a number of ways to significantly increase the reliability of the measures in Koplenig *et al.* (2017) and get even more meaningful results.

Yet another perspective

- There are probably a number of ways to significantly increase the reliability of the measures in Koplenig *et al.* (2017) and get even more meaningful results.
- But some limitations will remain, for example:

Yet another perspective

- There are probably a number of ways to significantly increase the reliability of the measures in Koplenig *et al.* (2017) and get even more meaningful results.
- But some limitations will remain, for example:
  - Bible translations are not very representative of a language, especially when it comes to syntax (cf. Maas, 2009)

Yet another perspective

- There are probably a number of ways to significantly increase the reliability of the measures in Koplenig *et al.* (2017) and get even more meaningful results.
- But some limitations will remain, for example:
  - Bible translations are not very representative of a language, especially when it comes to syntax (cf. Maas, 2009)
  - The measure relies much on orthographic word boundaries, which are not assigned consistently across languages.

Yet another perspective

- There are probably a number of ways to significantly increase the reliability of the measures in Koplenig *et al.* (2017) and get even more meaningful results.
- But some limitations will remain, for example:
  - Bible translations are not very representative of a language, especially when it comes to syntax (cf. Maas, 2009)
  - The measure relies much on orthographic word boundaries, which are not assigned consistently across languages.
- To complement studies based on parallel bible corpora, it would therefore be important to consider different data types as well.

The significance of data from language documentation

In particular, I would like to highlight the potential significance of data from language documentation.

- Language documentation data typically represents endangered languages.

The significance of data from language documentation

In particular, I would like to highlight the potential significance of data from language documentation.

- Language documentation data typically represents endangered languages.
- Typically, they are spoken in increasingly rare socio-linguistic settings that typologists believe to be conductive to special types of properties.

The significance of data from language documentation

In particular, I would like to highlight the potential significance of data from language documentation.

- Language documentation data typically represents endangered languages.
- Typically, they are spoken in increasingly rare socio-linguistic settings that typologists believe to be conductive to special types of properties.
- They are typically transcribed from natural discourse.

The significance of data from language documentation

In particular, I would like to highlight the potential significance of data from language documentation.

- Language documentation data typically represents endangered languages.
- Typically, they are spoken in increasingly rare socio-linguistic settings that typologists believe to be conductive to special types of properties.
- They are typically transcribed from natural discourse.
- In many cases, corpora from language documentation are small, but richly annotated.

Annotation layers in language documentation data

```
\ref .0009
\pt JI
\tx er      kevene, vyanten kevene yat        warsyosi
\mb er      kevene, vyanten kevene ya   -t    warsyosi
\ge 1P.IN every   man    every  3P   -DIST revere
\ps pron    q      n      q     agr -tam v

\ELANEnd 72.894
\ELANBegin 69.399
\ft we all, everyone used to respect him
^^I
```

Using annotations to measure complexity

- Despite their small size, richly annotated corpora from language documentation could still yield meaningful complexity measures for different layers:

Using annotations to measure complexity

- Despite their small size, richly annotated corpora from language documentation could still yield meaningful complexity measures for different layers:
- Morphological complexity could be measured by counting

Using annotations to measure complexity

- Despite their small size, richly annotated corpora from language documentation could still yield meaningful complexity measures for different layers:
- Morphological complexity could be measured by counting
  - the average/ maximum number of morphemes per word (degree of synthesis)

Using annotations to measure complexity

- Despite their small size, richly annotated corpora from language documentation could still yield meaningful complexity measures for different layers:
- Morphological complexity could be measured by counting
  - the average/ maximum number of morphemes per word (degree of synthesis)
  - the average/ maximum number of categories per morpheme/ word (degree of fusion)

Using annotations to measure complexity

- Despite their small size, richly annotated corpora from language documentation could still yield meaningful complexity measures for different layers:
- Morphological complexity could be measured by counting
  - the average/ maximum number of morphemes per word (degree of synthesis)
  - the average/ maximum number of categories per morpheme/ word (degree of fusion)
- Syntactic complexity could be measured by assessing entropy at the POS-level.

Sketching a way forward

To get a better handle on cross-linguistic variation in complexity, we need to:

- Compile a more complete survey of specific predictions from the typological (and other) literature.

Sketching a way forward

To get a better handle on cross-linguistic variation in complexity, we need to:

- Compile a more complete survey of specific predictions from the typological (and other) literature.
- Develop more fine-grained methods to explore complexity measures and their correlates in the parallel bible corpus.

Sketching a way forward

To get a better handle on cross-linguistic variation in complexity, we need to:

- Compile a more complete survey of specific predictions from the typological (and other) literature.
- Develop more fine-grained methods to explore complexity measures and their correlates in the parallel bible corpus.
- Explore other data types, in particular, richly annotated data from endangered languages, to consolidate and better understand measures and correlates.

References I

Atkinson, Quentin D. 2011. Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science*, **332**, 346–359.

Bisang, Walter. 2009. On the evolution of complexity: sometimes less is more in East and mainland Southeast Asia. *Pages 34–49 of:* Sampson, Geoffrey, Gil, David, & Trudgill, Peter (eds), *Language complexity as an evolving variable*. Oxford, New York: Oxford University Press.

Dryer, Matthew S. 2013. Order of subject, object and verb. *In:* Dryer, Matthew S., & Haspelmath, Martin (eds), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Ehret, Katharina, & Szmrecsanyi, Benedikt. 2016. An information-theoretic appraoch to assess linguistic complexity. *Pages 71–94 of:* Baechler, Raffaela, & Seiler, Guido (eds), *Complexity, isolation and variation*. Berlin, New York: de Gruyter.

References II

Everett, Daniel. 2005. Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology*, **46**, 621–646.

Futrell, Richard, Mahowald, Kyle, & Gibson, Edward. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the national academy of sciences*, **112**(33), 10336–10341.

Gildea, Daniel, & Jaeger, T. Florian. 2016. Human languages order information efficiently. *Proceedings of the National Academy of Sciences*, **preprint**.

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry*, **15**, 531–574.

Joseph, John E., & Newmeyer, Frederick J. 2012. All languages are equally complex: The rise and fall of a consensus. *Historiographia linguistica*, **39**, 341–368.

Juola, Patrick. 1998. Measuring linguistic complexity: The morphological tier. *Journal of Quantitative Linguistics*, **5**, 206–250.

References III

Juola, Patrick. 2008. Assessing linguistic complexity. *Pages 89–108 of:* Miestamo, Matti, Sinnemäki, Kaius, & Karlsson, Fred (eds), *Language complexity – typology, contact, change*. Studies in Language Companion Series, vol. 94. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Kilarski, Marcin. 2014. Complexity in the history of language study. *Poznan Studies in Contemporary Linguistics*, **50**(2), 157–168.

Kilarski, Marcin, & Dziubalska-Kołaczyk. 2012. On extremes in linguistic complexity. phonetic accounts of Iroquoian, Polynesian and Khoesan. *Historiographia Linguistica*, **39**(2/3), 279–303.

Koplenig, Alexander, Meyer, Peter, Wolfer, Sascha, & Müller-Spitzer, Carolin. 2017. The statistical trade-off between word order and word structure – large-scale evidence for the principle of least effort. *PLoS one*, **12**(3).

References IV

Maas, Utz. 2009. Orality versus literacy as a dimension of complexity. *Pages 164–177 of:* Sampson, Geoffrey, Gil, David, & Trudgill, Peter (eds), *Language complexity as an evolving variable*. Oxford, New York: Oxford University Press.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. *Pages 71–95 of: Consonant types and tone*. Southern California Occasional Papers in Linguistics, no. 1. Los Angeles: University of Southern California.

Mayer, Thomas, & Cysouw, Michael. 2014. Creating a massively parallel bible corpus. *Pages 26–31 of:* Choukri, K., T., Declerck, H., Loftsson, Maegaard, Bl, & Mariani, J. (eds), *Proceedings of the ninth international conference on language resources and evaluation*. Reykjavik, Iceland: European Language Resources Association (ELRA).

McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology*, **5**, 125–166.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago and London: University of Chicago Press.

References V

Pellegrino, François, Coupé, Christophe, & Marsico, Egidio. 2011. A cross-language perspective on speech information rate. *Language*, **87**, 539–558.

Riddle, Elizabeth M. 2008. Complexity in isolating languages: Lexical elaboration versus grammatical economy. *Pages 133–153 of:* Miestamo, Matti, Sinnemäki, Kaius, & Karlsson, Fred (eds), *Language complexity – typology, contact, change*. Studies in Language Companion Series, vol. 94. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Rijkhoff, Jan, & Bakker, Dik. 1998. Language sampling. *Linguistic Typology*, **2**, 263–314.

Sampson, Geoffrey. 2009. A linguistic axiom challenged. *Pages 1–18 of:* Sampson, Geoffrey, Gil, David, & Trudgill, Peter (eds), *Language complexity as an evolving variable*. Oxford, New York: Oxford University Press.

References VI

Siewierska, Anna. 1998. Variation in major constituent order: A global and a European perspective. *Pages 475–551 of:* Siewierska, Anna (ed), *Constituent order in the languages of Europe: Empirical approaches to language typology*. Berlin: Mouton de Gruyter.

Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford University Press.

Trade-offs between levels

Many researchers continue to assume certain complexity trade-offs between different structural levels (cf. Joseph & Newmeyer, 2012), for example:

- explicitly encoded information vs. reliance on pragmatics (*hot* vs. *cool* languages) (Huang, 1984)

Trade-offs between levels

Many researchers continue to assume certain complexity trade-offs between different structural levels (cf. Joseph & Newmeyer, 2012), for example:

- explicitly encoded information vs. reliance on pragmatics (*hot* vs. *cool* languages) (Huang, 1984)
- Rich case structure vs. flexible word order (Siewierska, 1998)

Trade-offs between levels

Many researchers continue to assume certain complexity trade-offs between different structural levels (cf. Joseph & Newmeyer, 2012), for example:

- explicitly encoded information vs. reliance on pragmatics (*hot* vs. *cool* languages) (Huang, 1984)
- Rich case structure vs. flexible word order (Siewierska, 1998)
- complex syllable structure correlates with low tonal complexity (Matisoff, 1973)

Trade-offs between levels

Many researchers continue to assume certain complexity trade-offs between different structural levels (cf. Joseph & Newmeyer, 2012), for example:

- explicitly encoded information vs. reliance on pragmatics (*hot* vs. *cool* languages) (Huang, 1984)
- Rich case structure vs. flexible word order (Siewierska, 1998)
- complex syllable structure correlates with low tonal complexity (Matisoff, 1973)
- isolating morphology correlates with a rich inventory of processes at the lexicon-syntax interface (Bisang, 2009; Riddle, 2008);

Equal complexity pro and con

- Notions about *primitive* as opposed to *complex* languages were part of the racist and colonialist discourse until the 20th century (cf. Kilarski & Dziubalska-Kołaczyk, 2012; Kilarski, 2014).

Equal complexity pro and con

- Notions about *primitive* as opposed to *complex* languages were part of the racist and colonialist discourse until the 20th century (cf. Kilarski & Dziubalska-Kołaczyk, 2012; Kilarski, 2014).
- Linguistic text books from later decades uniformly asserted that there is no basis for such a distinction and that all languages are more or less equally complex.

Equal complexity pro and con

- Notions about *primitive* as opposed to *complex* languages were part of the racist and colonialist discourse until the 20th century (cf. Kilarski & Dziubalska-Kołaczyk, 2012; Kilarski, 2014).
- Linguistic text books from later decades uniformly asserted that there is no basis for such a distinction and that all languages are more or less equally complex.
- Since the 1980s, linguists increasingly challenge the equal-complexity claim.

Claims against the equal-complexity hypothesis

- Nichols (1992); Trudgill (2011): (certain types of) linguistic complexity tend to accumulate under specific socio-linguistic conditions.

Claims against the equal-complexity hypothesis

- Nichols (1992); Trudgill (2011): (certain types of) linguistic complexity tend to accumulate under specific socio-linguistic conditions.
- McWhorter (2001): Creole languages are radically simpler than non-creole languages.

Claims against the equal-complexity hypothesis

- Nichols (1992); Trudgill (2011): (certain types of) linguistic complexity tend to accumulate under specific socio-linguistic conditions.
- McWhorter (2001): Creole languages are radically simpler than non-creole languages.

Claims against the equal-complexity hypothesis

- Nichols (1992); Trudgill (2011): (certain types of) linguistic complexity tend to accumulate under specific socio-linguistic conditions.
- McWhorter (2001): Creole languages are radically simpler than non-creole languages.

Sampson (2009)

There cannot be many current topics of academic debate which have greater general human importance than this one [i.e. cross-linguistic variation in complexity].

Manipulations to target specific structural levels

- Juola (1998) replaces (type-wise) each non-blank character string by a randomly chosen number, then measures differences between compressed original and compressed degraded texts to estimate morphological complexity.

Manipulations to target specific structural levels

- Juola (1998) replaces (type-wise) each non-blank character string by a randomly chosen number, then measures differences between compressed original and compressed degraded texts to estimate morphological complexity.
- Ehret & Szmrecsanyi (2016) randomly delete 10% of characters to degrade morphological structure; syntactic structure is degraded by a random deletion of 10% of word tokens.