Systematicity
Why everything works
stylometry

ESU 24 @ Cluj-Napoca
Jeremi Ochab, Artjoms Šeļa

manga: Dragon Ball

Reconsidering Sausurre

- Iconic versus symbolic signs

- Iconicity is <u>also</u> conventional

- However, all form in language is non-arbitrary:
  systematicity of language

Arbitrariness, Iconicity, and Systematicity in Language

Mark Dingemanse [1], Damián E Blasi [2], Gary Lupyan [3], Morten H Christiansen [4], Padraic Monaghan [5]

How arbitrary is language?

Padraic Monaghan ✉, Richard C. Shillcock, Morten H. Christiansen and Simon Kirby
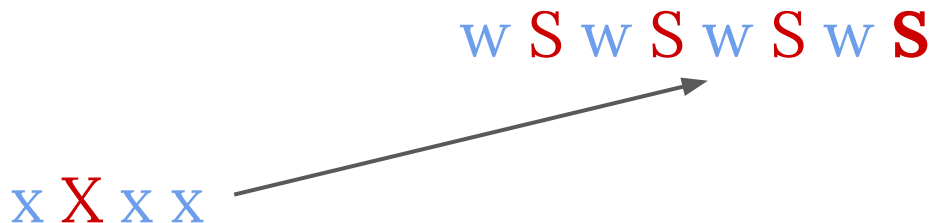
Reconsidering Sausurre

- Iconic versus symbolic signs

- Iconicity is <u>also</u> conventional

- However, all form in language is non-arbitrary:
  systematicity of language

manga: Dragon Ball

w S w S w S w **S**

x X x x

Fitting **4-syllable** word to a line of RU iambic tetrameter

w S w S w S w **S**

x X x x

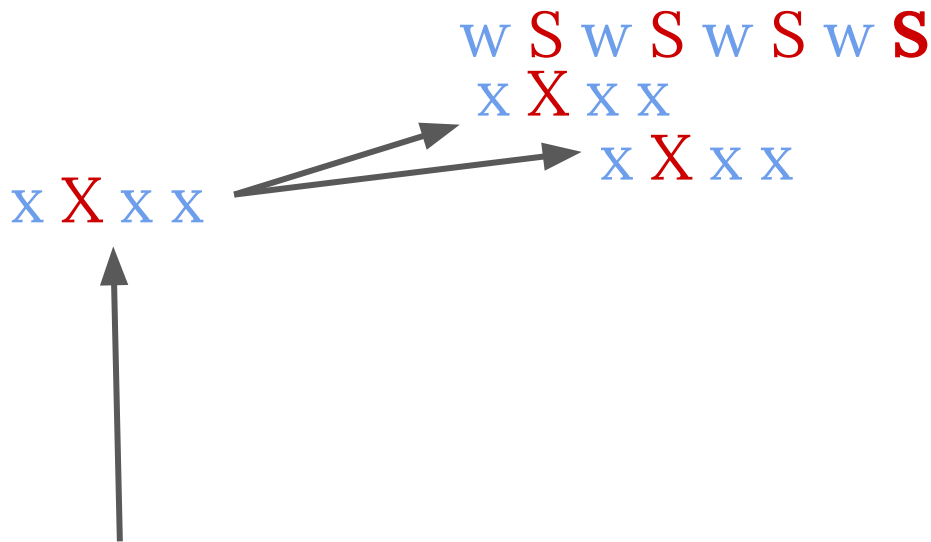Longer words: ADJ or VERB
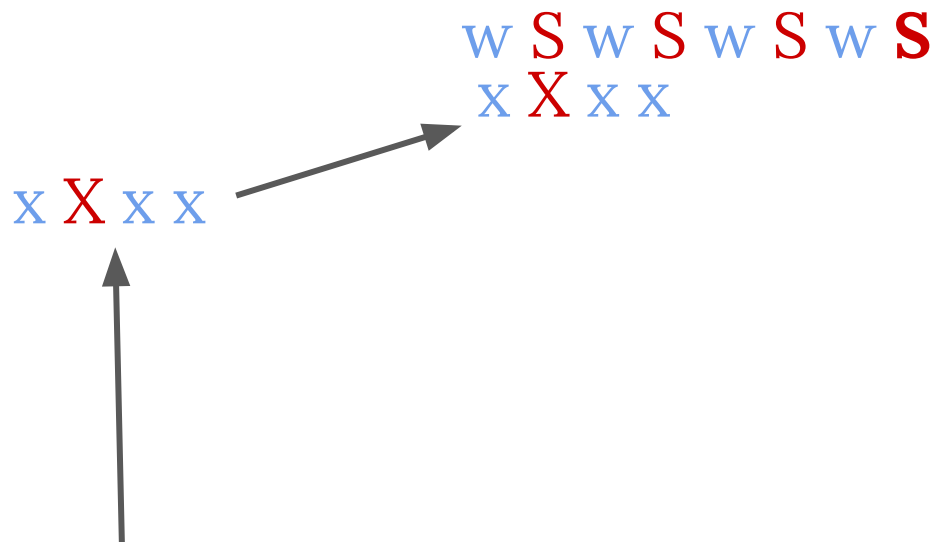
xXxx: twice as likely to be ADJ in RU

w S w S w S w **S**

x X x x

x X x x

Longer words: ADJ or VERB

xXxx: twice as likely to be ADJ in RU

Linguistics of verse

w S w S w S w **S**
x X x x
x X x x
x X x x

Longer words: ADJ or VERB

xXxx: twice as likely to be ADJ in RU

Linguistics of verse

w S w S w S w **S**

x X x x

x X x x

x X x x

ADJ preference

Longer words: ADJ or VERB

xXxx: twice as likely to be ADJ in RU

Linguistics of verse

ADJ(4) . NOUN(2)

w S w S w S w **S**
x X x x
x X x x

x X x x

ADJ preference
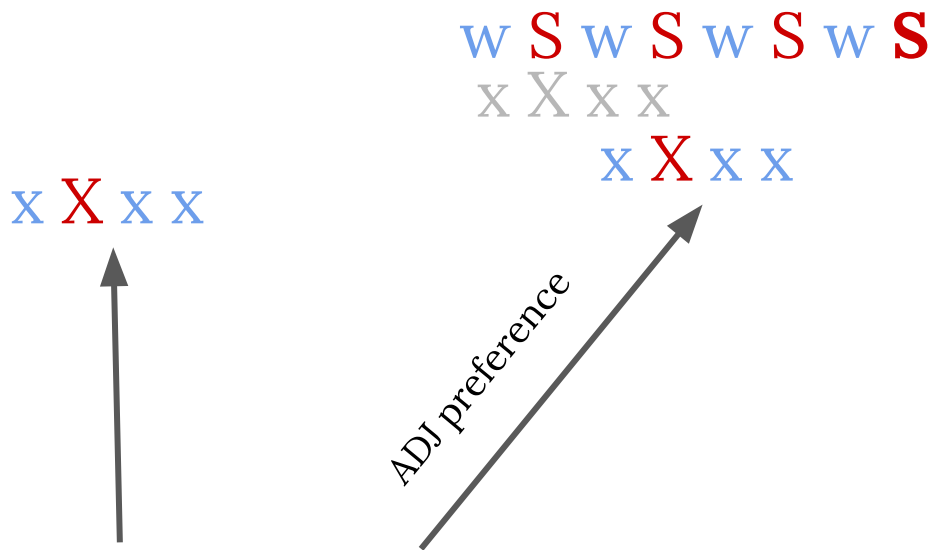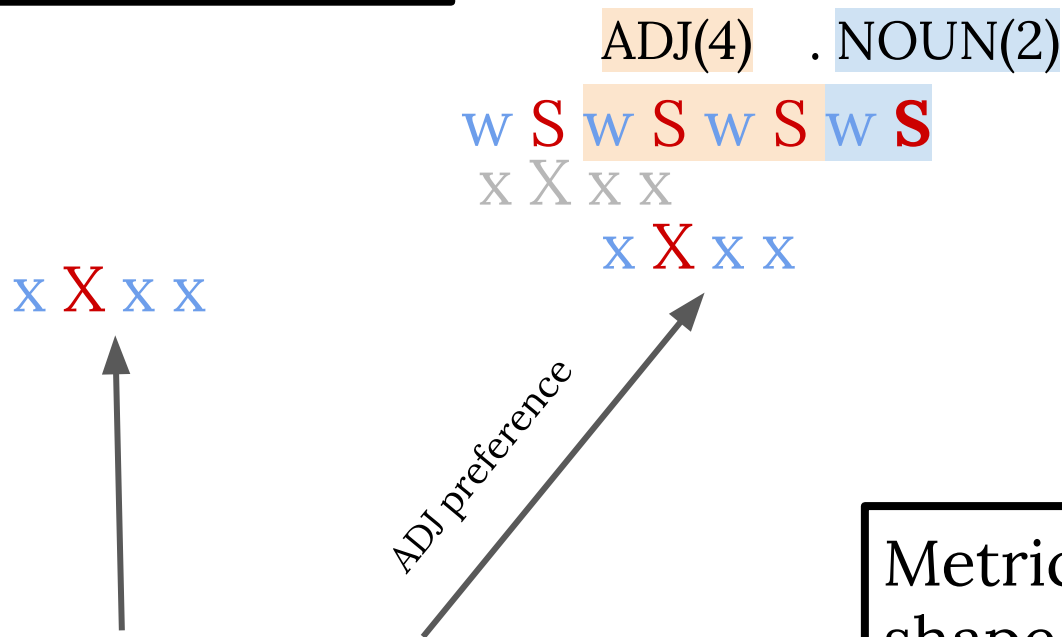
Metrical constraints shape formulas

Longer words: ADJ or VERB

xXxx: twice as likely to be ADJ in RU

Confounding differences: morphology & syntax

Once upon a midnight dreary, while I pondered, weak and weary

POS + word syllable lengths

ADV_1 ADP_2 DET_1 NOUN_2 ADJ_2, SCONJ_1 PRON_1 VERB_2, ADJ_1 CCONJ_1 ADJ_2

Confounding differences: morphology & syntax

POS + length 2-grams:

ADV_1 ADP_2, ADP_2 DET_1,
DET_1 NOUN_2, NOUN_2 ADJ_2...

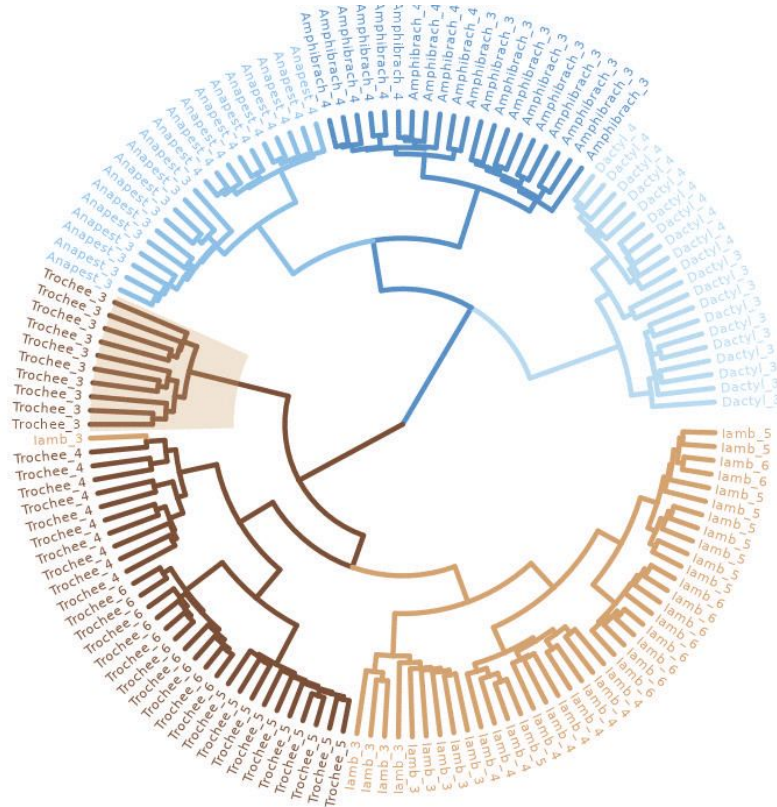18-19 c. Russian poetry, NRC

100 lines per sample
10 samples per meter
14 forms (with feet-length variations)
Jensen-Shannon Divergence + Ward's clustering

WARM hue: binary meters (iambs, trochees)

COLD hue: ternary meters (anapests etc.)

- There are no independent levels in language

- To some extent: each level can be predicted using another

- Word frequencies are never "just" word frequencies

- Function words <- grammar templates <- idiolect <- register <- intention <- society / culture