On the relation between word length and phoneme frequencies

Ján Mačutek, Radek Čech and Michaela Koščová

CILC 2025, Salamanca

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

Financované Európskou úniou NextGenerationEU

PLÁN OBNOVY

ÚRAD PODPREDSEDU VLÁDY SLOVENSKEJ REPUBLIKY PRE PLÁN OBNOVY A ZNALOSTNÚ EKONOMIKU

Summary

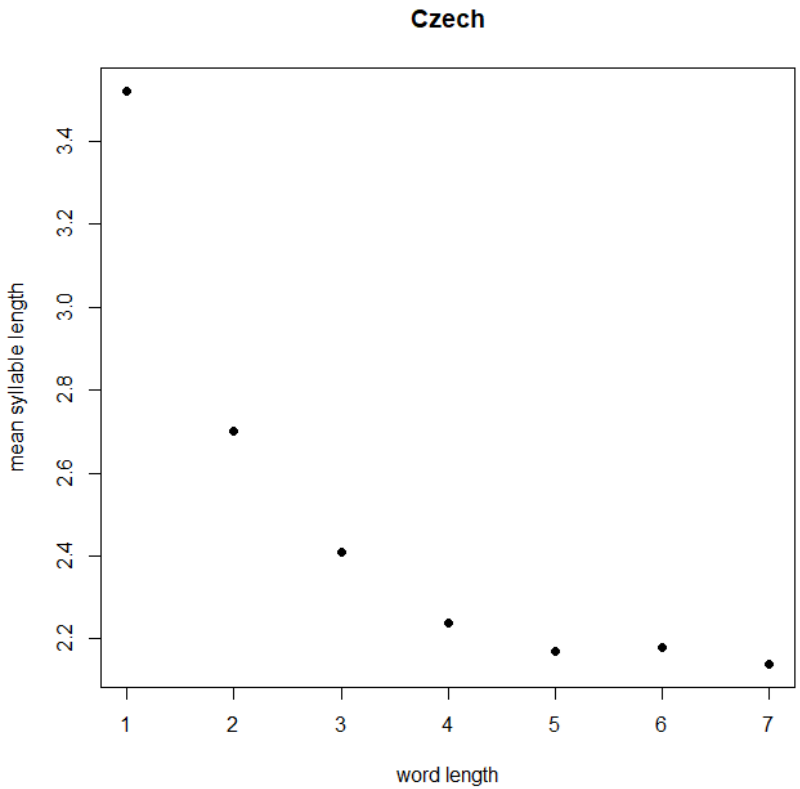
Introduction

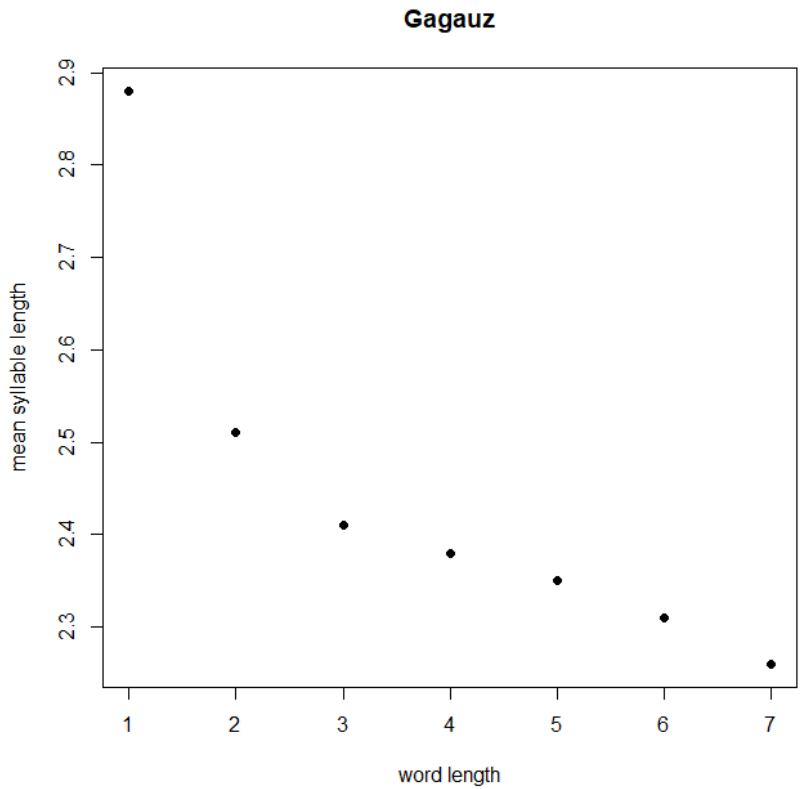
- principle of least effort
  - George K. Zipf
  - e.g. more frequent words are shorter
  - obviously, word tokens are considered

Menzerath-Altmann law

- principle of least effort
- Menzerath-Altmann law
  - the longer the language construct, the shorter its constituents (on average)
  - the longer the word, the shorter the mean syllable length
  - word length in syllables, syllable length in phonemes
  - valid for word types
  - principle of least effort shapes vocabulary
  - reducing effort (producing words, recalling them from memory)
  - very loosely speaking, I want to use "cheap" words – if I must say a words with many syllables, I shorten syllables

Menzerath-Altmann law for Czech & Gagauz





Menzerath-Altmann law is valid in all languages

We thought so...until recently.

Menzerath-Altmann law is valid in all languages...or not?

- languages with simple syllable structure (only V and CV syllables) do not obey the Menzerath-Altmann law
- validity of the Menzerath-Altmann law would lead to long vowel clusters
- a clash between the Menzerath-Altmann law and the *horror aequi* principle

Language material

Translations of the New Testament (27 books) into 6 languages from 5 different language families

- Bilua (Central Solomon, Solomon Islands, 8,700 speakers)
- Bola (Austronesian, Papua New Guinea, 14,000 speakers)
- Czech (Indo-European, Czech Republic, 12 million speakers)
- Gagauz (Turkic, Moldova, 150,000 speakers)
- Jamamadi (Arawan, Brazil, 1,000 speakers)
- Tongan (Austronesian, Tonga, 187,000 speakers)

Bilua, Bola, Jamamadi, Tongan – only CV and V syllables
all these languages have shalow orthographies – easy to determine the numer of syllables and phonemes in words

How is principle of least effort realized in vocabulary of these languages?

Sonority hierarchy

phonemes	sonority index
low vowels	10
mid vowels	9
high vowels and semivowels	8
rhotics	7
laterals	6
nasals	5
voiced fricatives	4
voiceless fricatives	3
voiced plosives and affricates	2
voiceless plosives and affricates	1

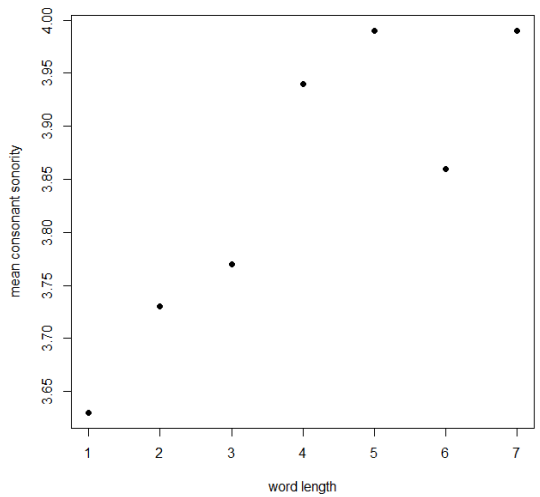
Back to languages obeying the Menzerath-Altmann law

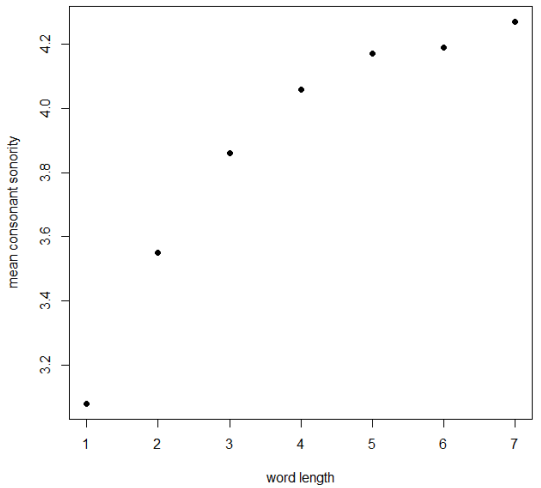
- consonant clusters are difficult to pronounce
- shortening of syllables reduces consonant clusters

Consequence of the Menzerath-Altmann law

- if syllables get shorter with increasing word lenth, the proportion of vowels increases in longer words
- vowels have higher sonority than consonants
- therefore, the mean phoneme sonority should increase

Czech & Gagauz



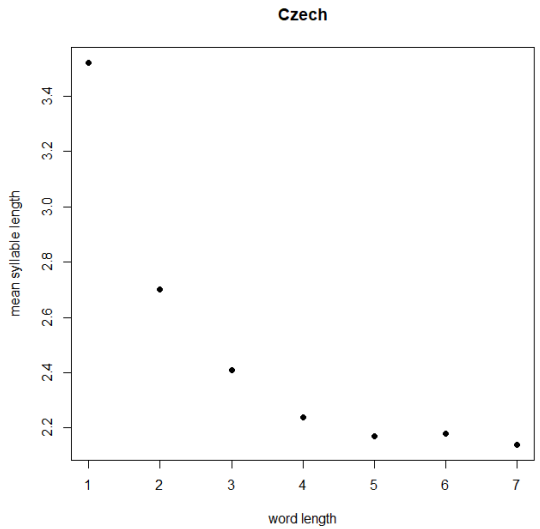


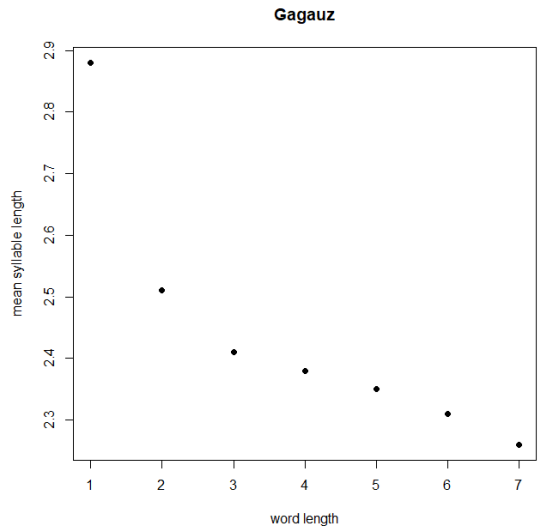
Czech & Gagauz

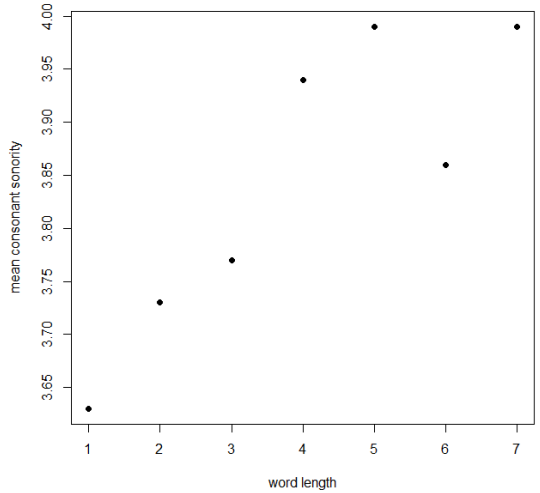
But there's more of effort saving...

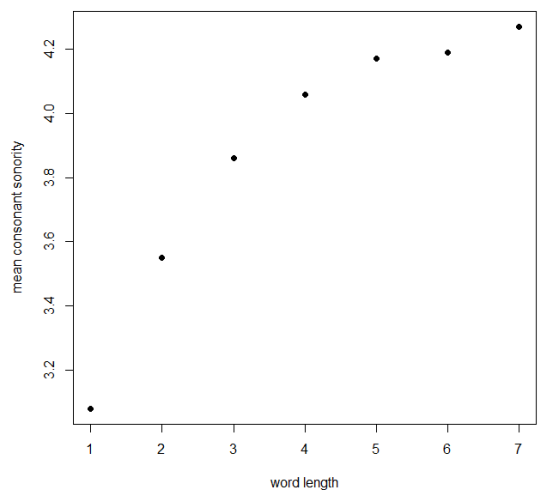
Not only syllables get shorter in longer words. also, consonants get more sonorous.
Thus, they get more "vowel-like" – lenition, "less dramatic" changes from vowels to consonants and back.

Czech & Gagauz



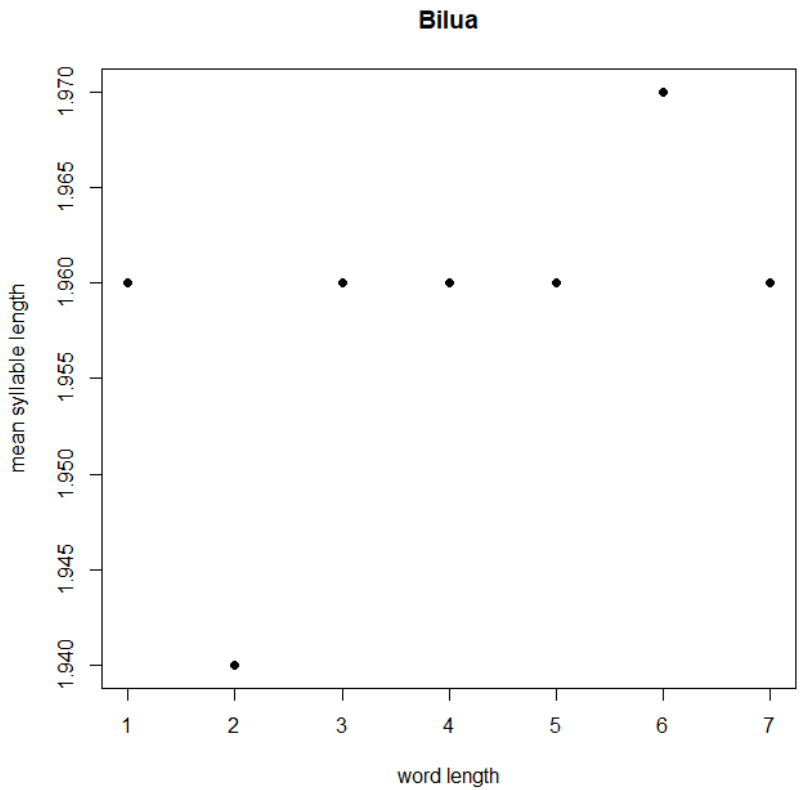






Bilua

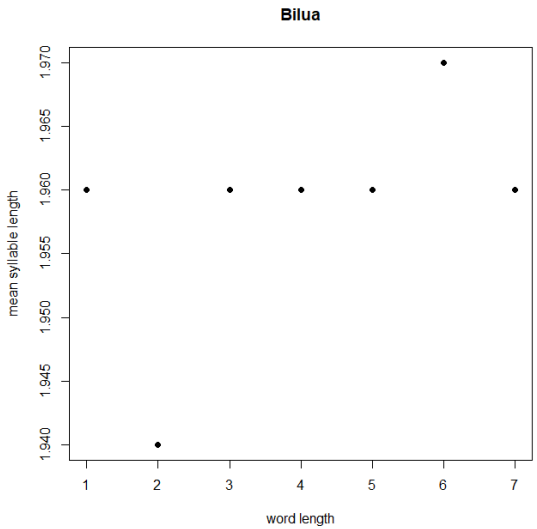
- first syllable can be V or CV, all other are CV
- no chance for a systematic decrease of the mean syllable length

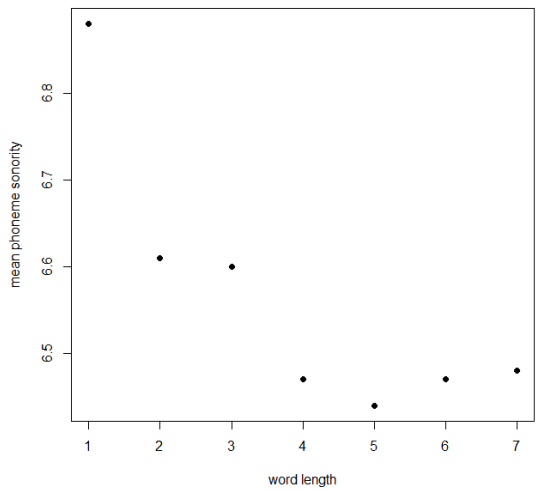


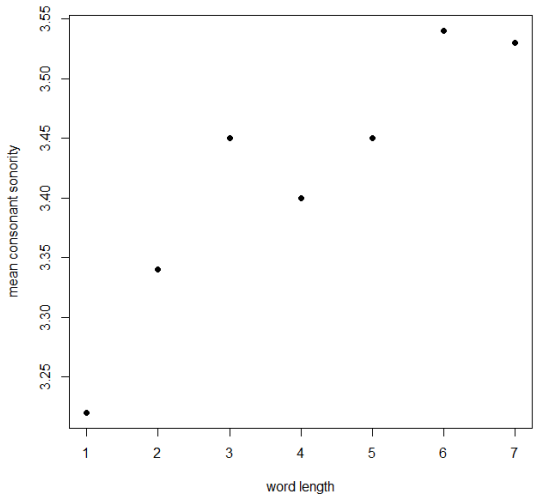
Bilua

- the mean phoneme sonority decreases (because the proportion of consonants increases, and consonants are less sonorous then vowels)
- the mean consonant sonority increases, for the same reason as in Czech and Gagauz

Bilua


Bilua
mean syllable length
1.940
1.945
1.950
1.955
1.960
1.965
1.970
1
2
3
4
5
6
7
word length

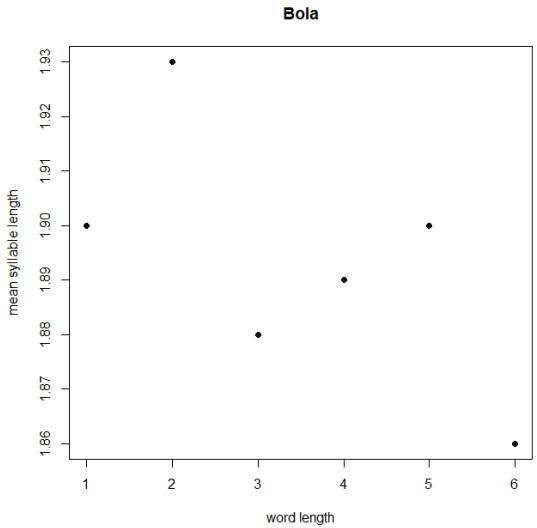

mean phoneme sonority
6.5
6.6
6.7
6.8
1
2
3
4
5
6
7
word length

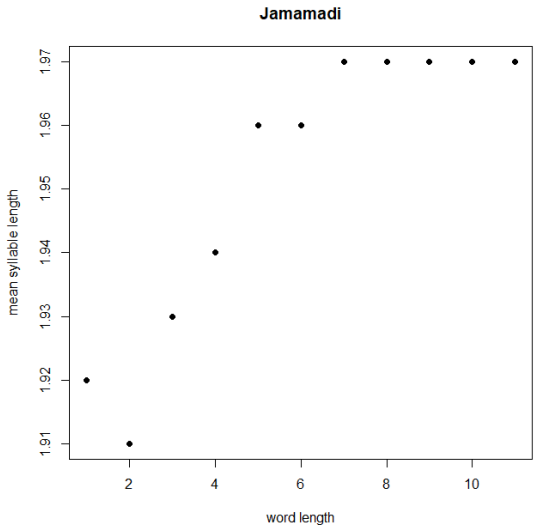

mean consonant sonority
3.25
3.30
3.35
3.40
3.45
3.50
3.55
1
2
3
4
5
6
7
word length

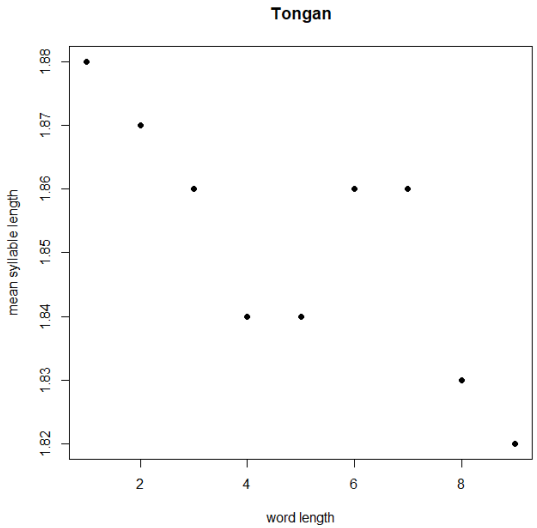
Bola, Jamamadi & Tongan

- only CV and V syllables (no positional restrictions like in Bilua)
- the mean syllable length cannot decrease with increasing word length (vowel clusters would be created)
- some languages can tolerate them to some extent, but they shoould not be too long

Bola, Jamamadi & Tongan



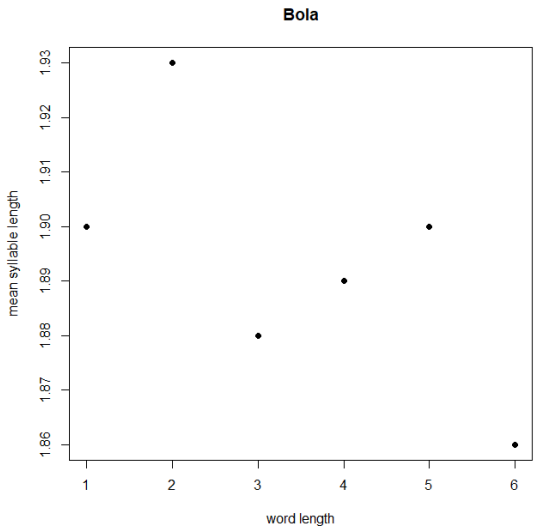


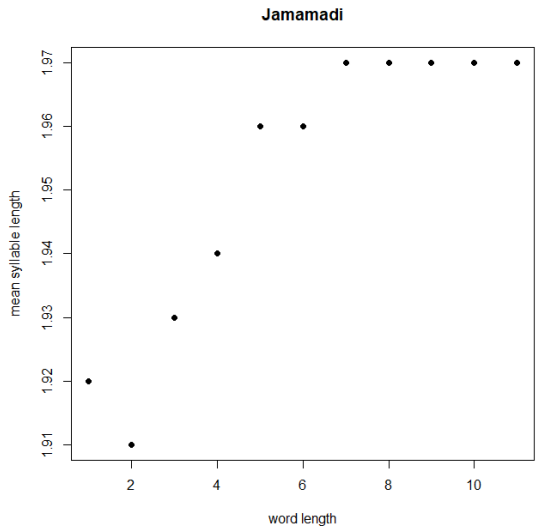


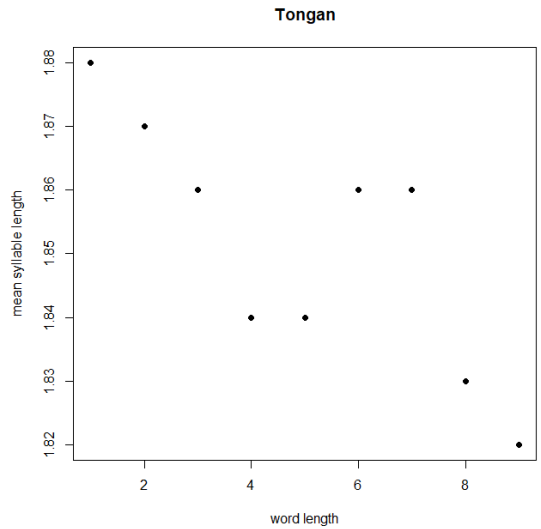
Bola, Jamamadi & Tongan

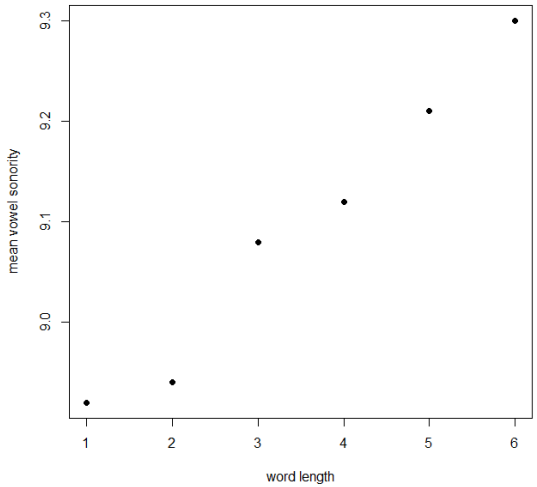
- low vowels are easier to pronounce (phonetics – high vowels require higher pressure of air, i.e. greaer effort)
- the mean vowel sonority increases with increasing word length in these langiages – they prefer easier to pronounce vowels

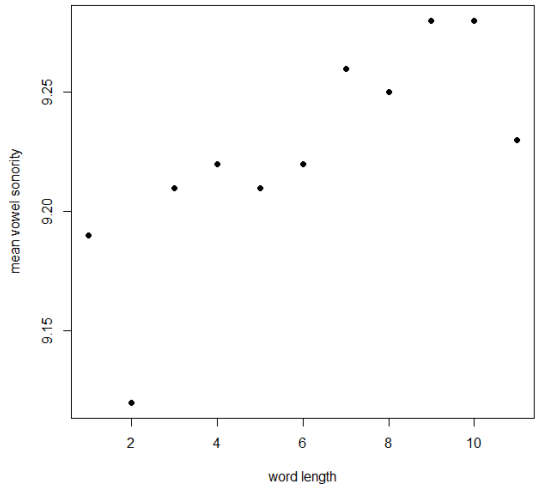
Bola, Jamamadi & Tongan

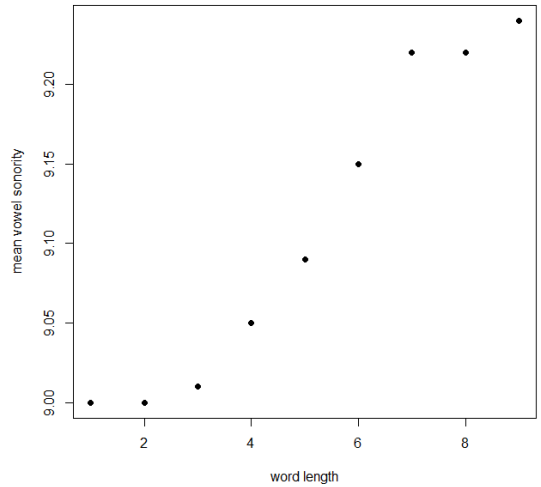






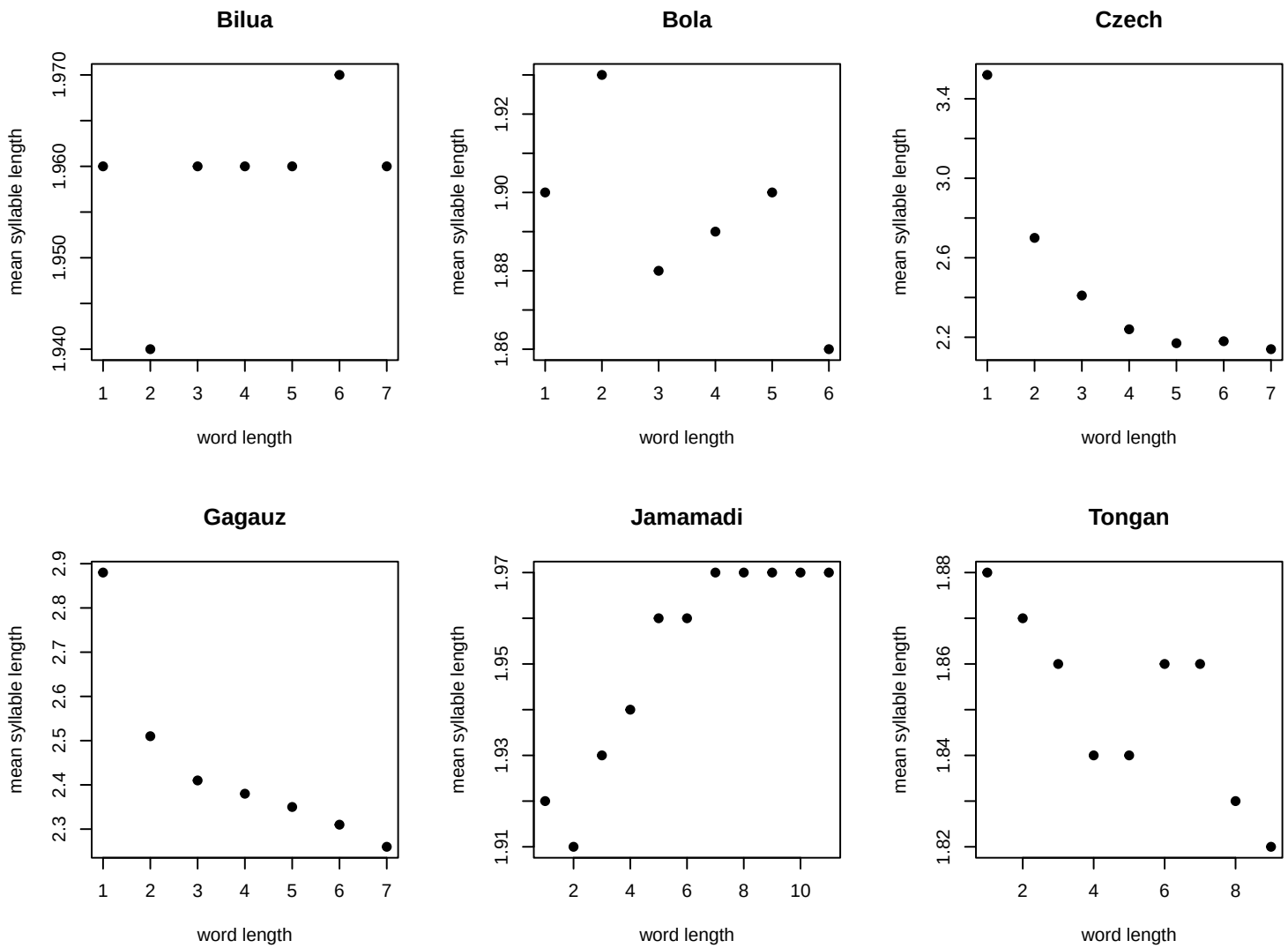






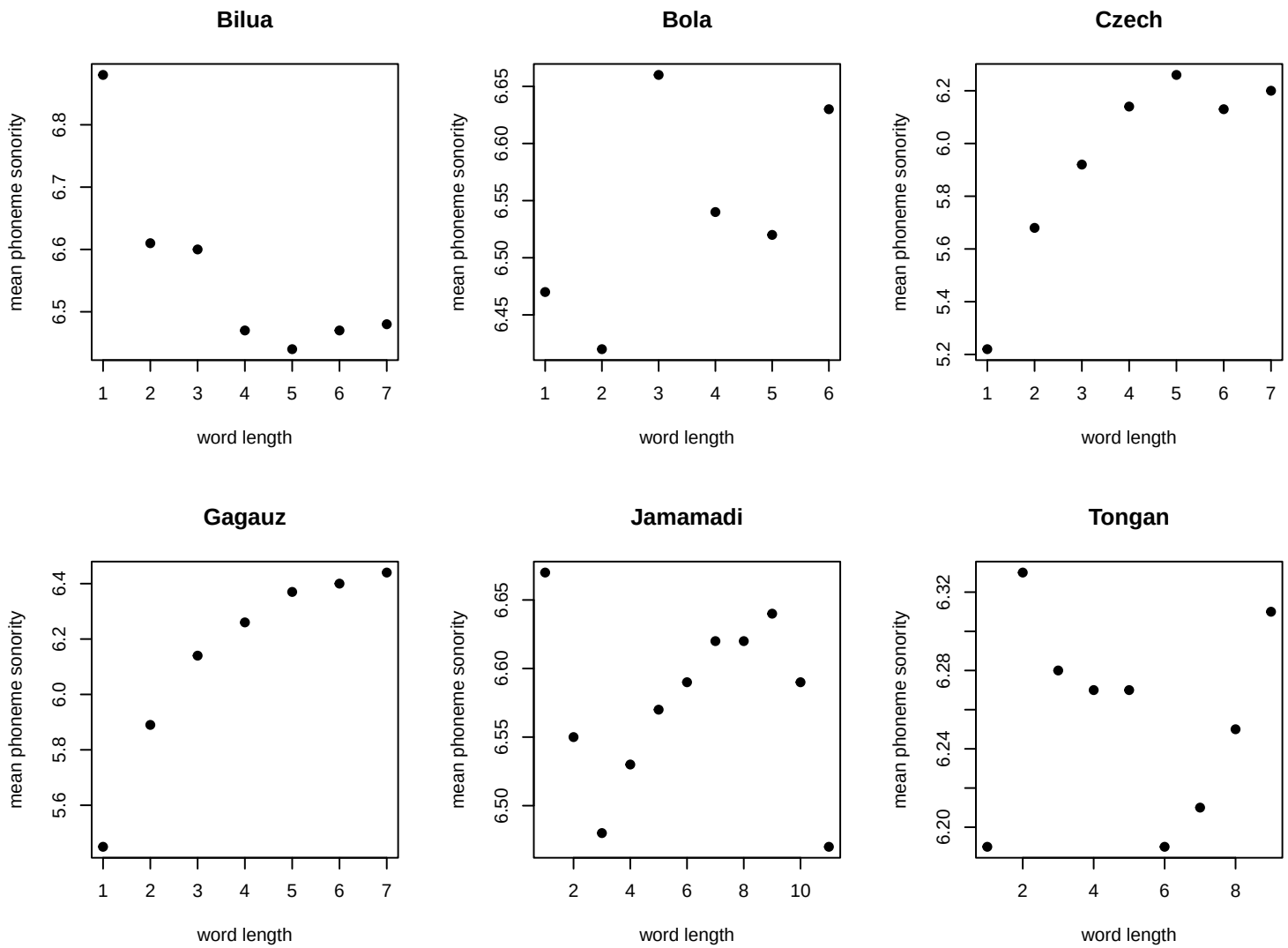
Summary of results

Menzerath-Altmann law



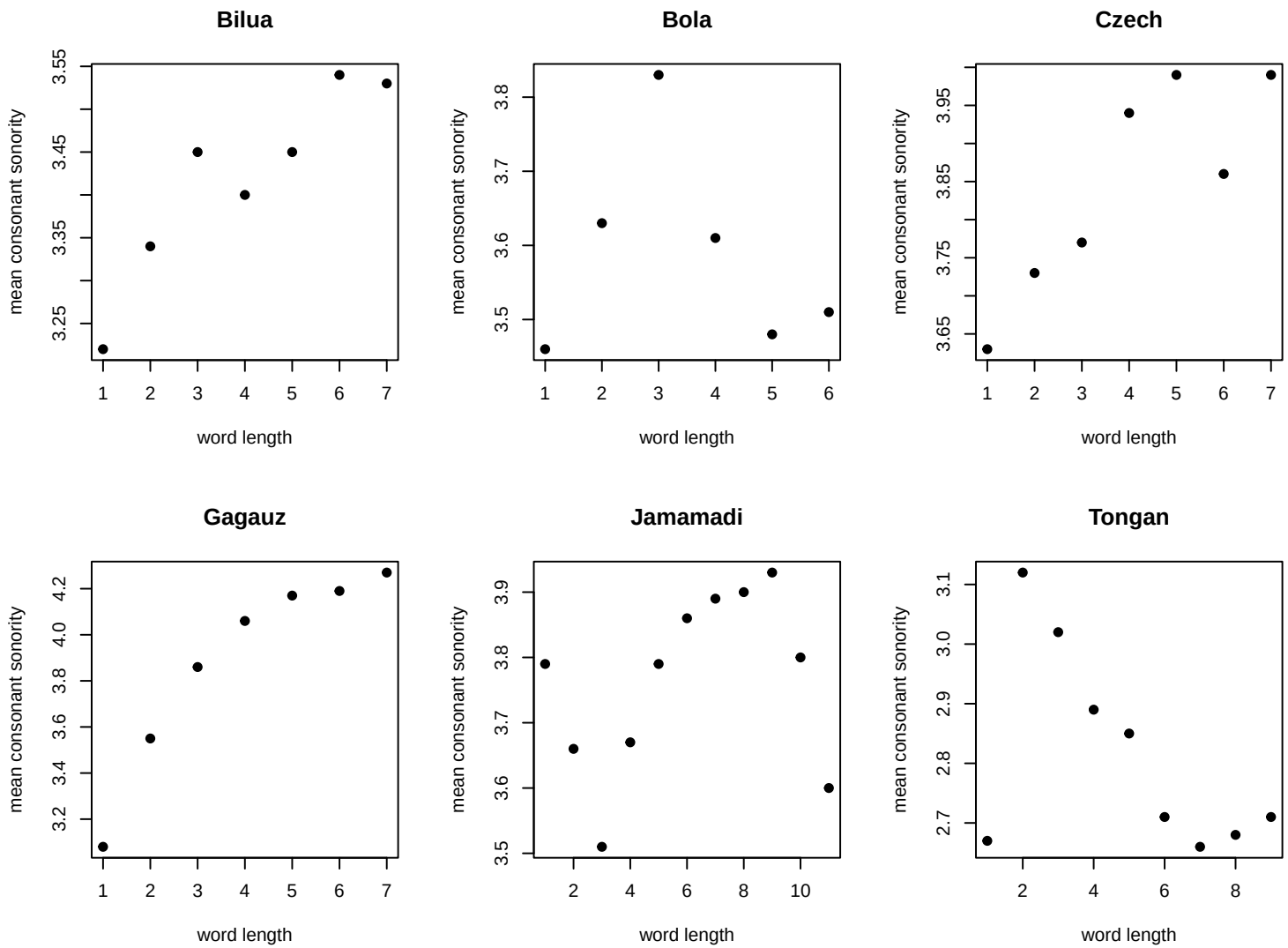
Relationship between word length and the mean syllable length

Mean phoneme sonority



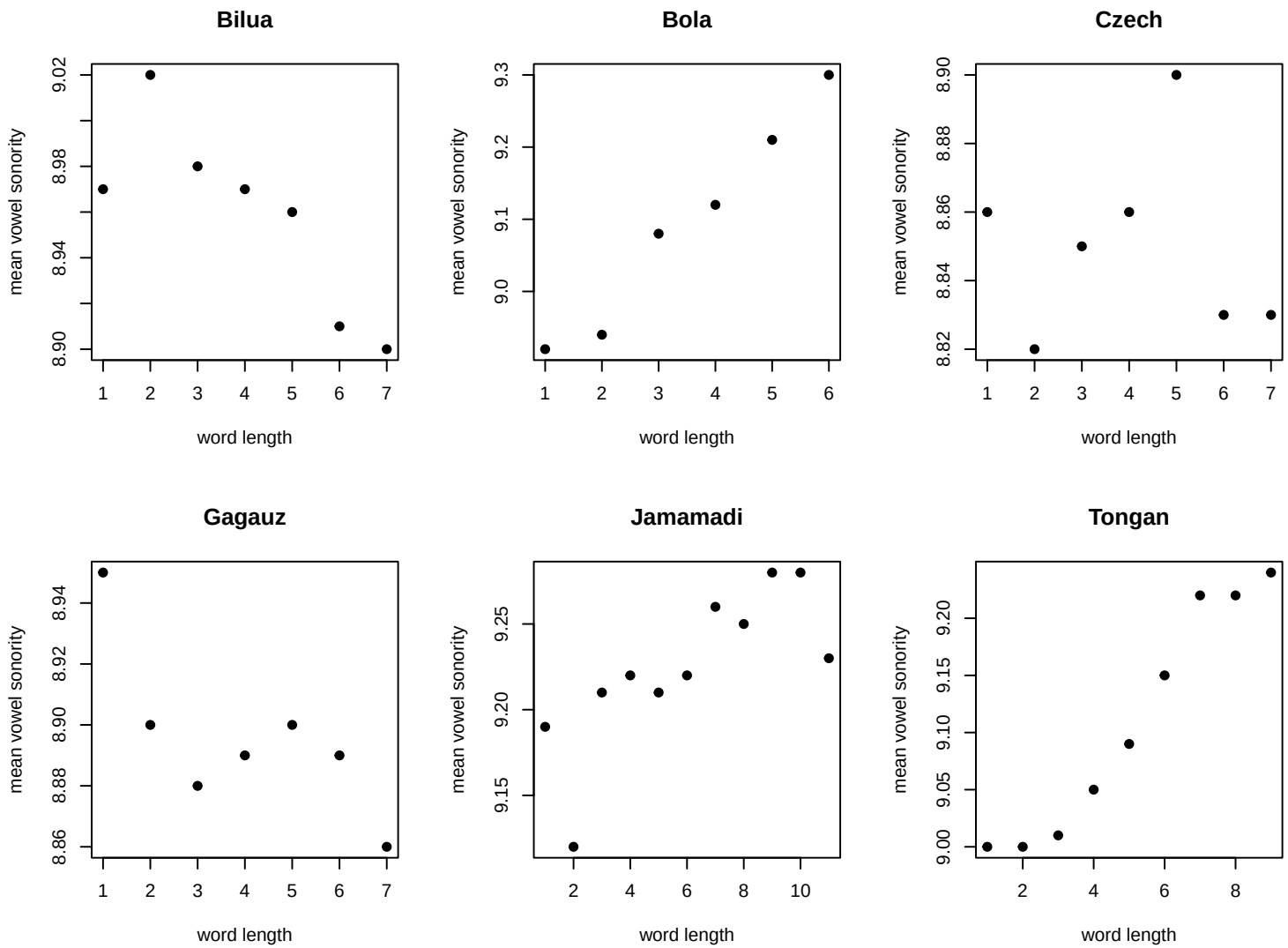
Relationship between word length and the mean phoneme sonority

Mean consonant sonority



Relationship between word length and the mean consonant sonority

Mean vowel sonority



Relationship between word length and the mean vowel sonority

Conclusion

The same principle (of least effort), but different realizations:

- if language can obey the Menzerath-Altmann law, greater effort conected to producing longer words is compensated by shorter syllables and consonant lenition (going from a consonant to a vowel and back becomes less demanding)

Conclusion

- if syllabic structure fights against the Menzerath-Altmann law, we observe two different strategies:
  - Bilua has another restriction, it increases sonority of its consonants in longer words
  - Bola, Jamamadi, and Tongan increase sonority of its vowels
  - both strategies cannot be applied simultaneously – it would lead to lower exploitation of high vowels and less sonorant consonants, and consequently a choice of phonemes in long words would be very limited with a limited phoneme inventory one faces the dilemma of either having more synonyms (more difficult to decode fro the hearer), or making words even longer (more difficult to produce for a speaker)

Zipf was right, languages find themselves in some (dynamic) equilibrium (between requirements of both speaker and hearer).

Gracias por su atención

Ján Mačutek

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00748.

Radek Čech

This work was funded by the European Regional Development Fund project "A lifetime with language: the nature and ontogeny of linguistic communication (LangInLife)" (reg. no.: CZ.02.01.01/00/23$_0$25/0008726

Michaela Koščová

Funded by VEGA 2/0120/24 Theoretical Properties and Applications of Special Classes of Probability Distributions.