

Bonito

Najděte místa, kde je v korpusu
chybně provedená segmentace na
věty

Věta

- Různé definice
 - jazyková jednotka, většinou charakterizovaná tím, že obsahuje přísudek (většinou sloveso)

Zdroje chyb

- Překlepy
- Linearizace dokumentu
- Konvence

Zkratky

- Skoro vždy správně
- Apod \.
 - Apod., Apod.)
- P. S.
 - P \. S \.

Jména

- Seznamy jmen
 - Zřejmě konvence
 -
- W \. A \. Mozart – OK

Různé

- Řadové číslovky
 - `\. [0-9]+ \.`
 - `„[0-9]+” „\.” [lemma=“leden”]`
- Jména
 - `[A-Z]+ \. [A-Z]+ \.`

Náhodné

- PhD \. , - 25/26