

Činnosti využívající morfologii

- morfologická analýza
- morfologické značkování
- částečná morfologická disambiguace

↳ Pokud ale byla syntaktická špatně,
tak ta disambiguace uřadila všechny značky
a tedy nezačal žádný rozumný význam.

↳ Používá se na klasifikaci dvýbní věty.
(pouze řekne, že věta je špatná)

- Lemmatizace

- proces výběru základního tvaru, ze kterého bylo slovo odvozeno

- Stemming

- odřezávání koncovek, než nám zůstane jen kořen
- blbě je, pokud má jazyk hodně koncových alternancí
- tudíž v češtině to moc nefunguje. (pro angličtinu je to ale super)

- Generování

- proces výběru správného slovního tvaru, pokud známe lemmu a jeho vzor.

„ý“ na konci kromě tří výjimek garantuje přidání jmeno (úřadý, prý, celý)

Lindat

- přehled všech morfologických / syntaktických / ... analýz

Inačkování

Chceme: $\hat{t}_1^n \approx \underset{t_1^n}{\operatorname{argmax}} \prod_{i=1}^n P(w_i | t_i) \cdot P(t_i | t_{i-1})$

↳ lze řešit pomocí HMM (hidden markov model)

Co lze dělat s HMM:

- 1) Rozpoznávání - snažíme se zjistit pravděpodobnost, že jde o nějakou konkrétní formu (mám špatnou formu SP2 a ptám se, jaká je past, že to je SP2.)
- 2) Dekódování - hledání nejpravděpodobnější posloupnosti sluchových stavů

Kontrola přichlepu

- 1) Musí být malý rozdíl mezi přichlepy
- 2) Systém nesmí dělat falešné výsledky
- 3) Slova, co systém uzná, nesmí být označena špatně
- 4) Čas na zpracování musí být velmi krátký
- 5) Uchytění musí být co nejvíce automatické

↳ Tyto podmínky je těžké všechny splnit.