

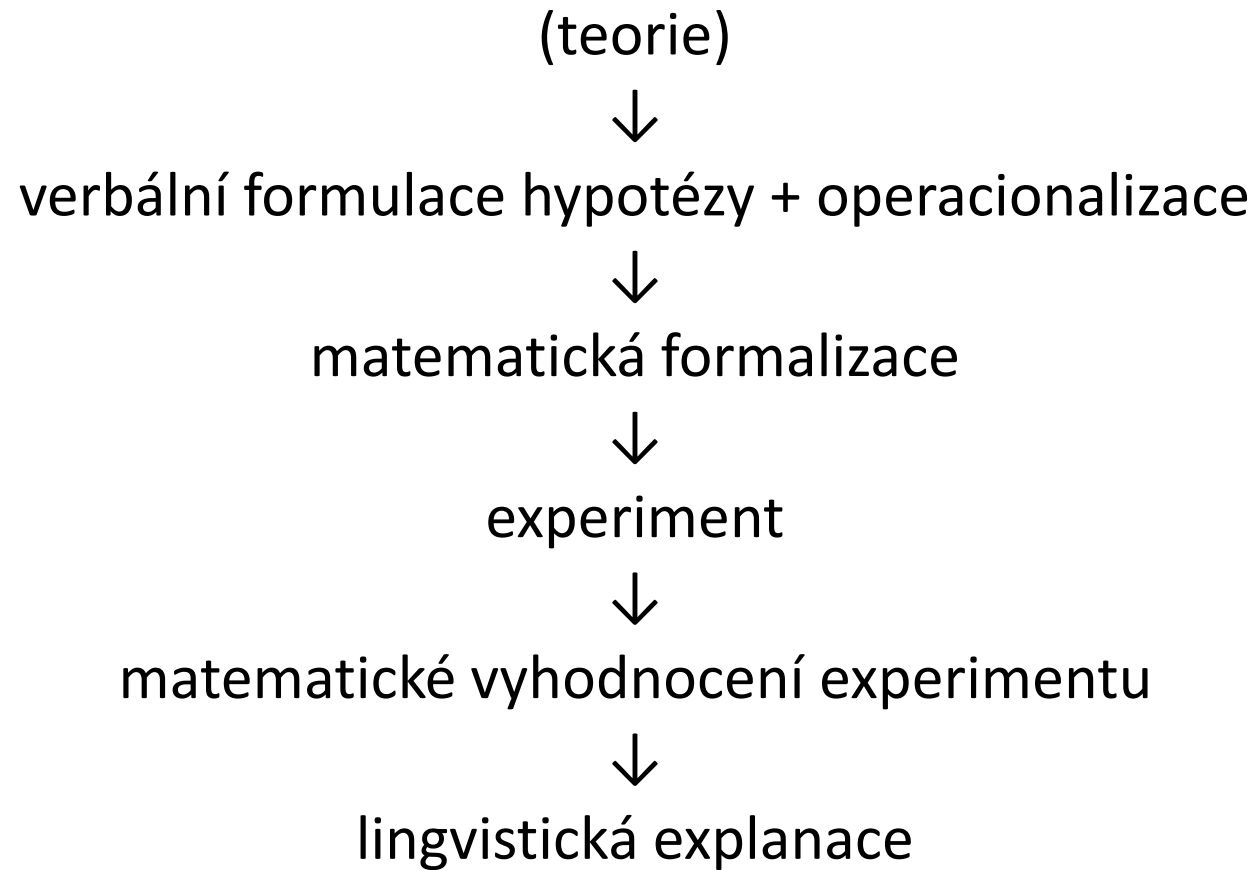
Úvod do kvantitativní lingvistiky

ZS 2024

Kvantitativní lingvistika

- co je to?

Metodologie



Kvantitativní lingvistika - perspektivy

- srov. např. Stefanowitsch, Anatol. (2020). *Corpus linguistics: A guide to the methodology*.
 - srov. kap.
 - 3.3 Hypotheses in context: The research cycle

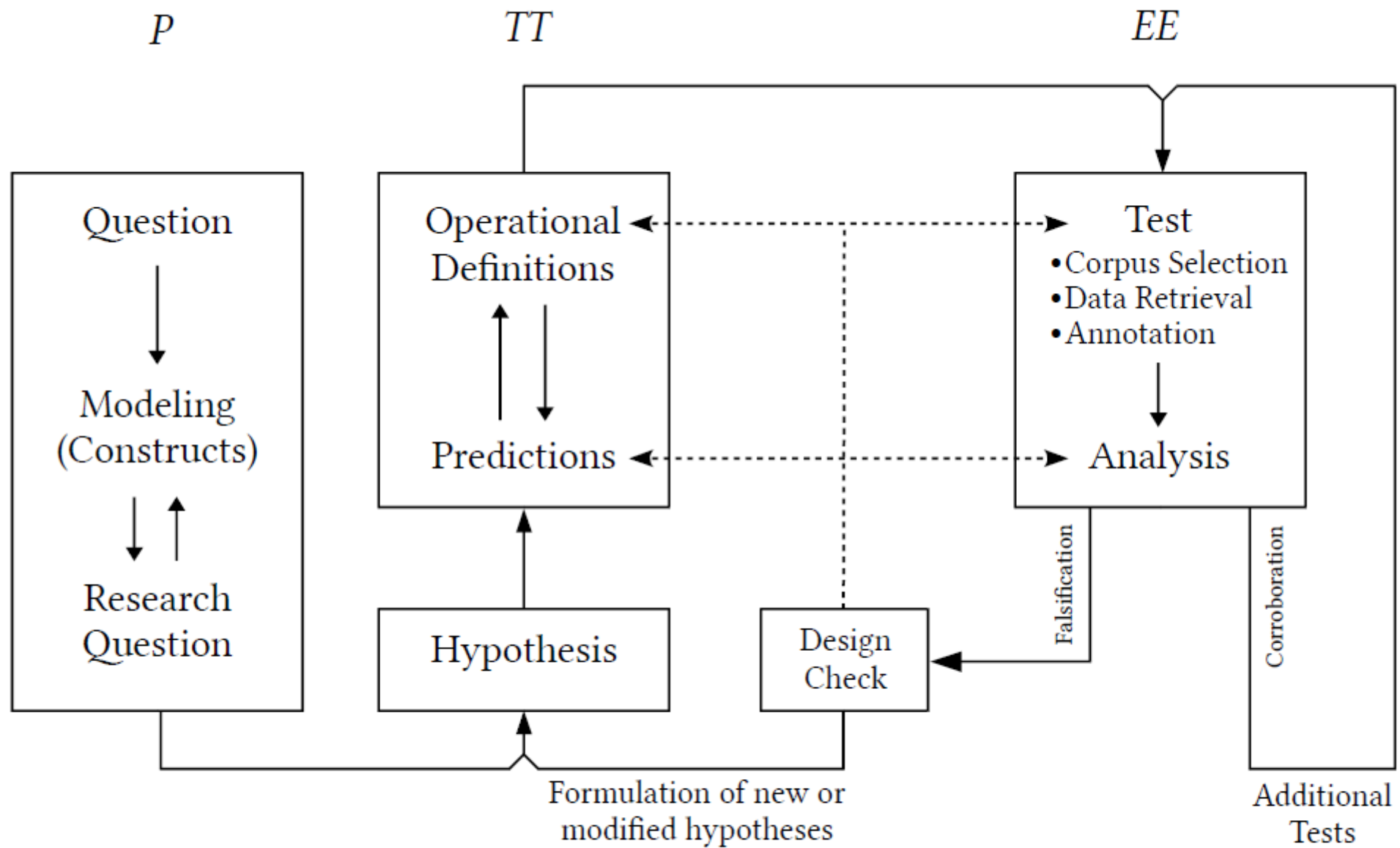


Figure 3.1: The scientific research cycle

Teorie

- Co je teorie?
- Jak poznat, která teorie je lepší než jiná?
- Plyne něco z teorie?
- Lingvistické teorie?

Teorie

- Co je teorie?
- Jak poznat, která teorie je lepší než jiná?
- Plyne něco z teorie?
- Lingvistické teorie?

- Wikipedie
 - teorie:
 - <https://cs.wikipedia.org/wiki/Teorie>
 - teorie (přírodní věda):
 - [https://cs.wikipedia.org/wiki/Teorie_\(p%C5%99%C3%ADrodn%C3%AD_v%C4%9Bda\)](https://cs.wikipedia.org/wiki/Teorie_(p%C5%99%C3%ADrodn%C3%AD_v%C4%9Bda))

Hypotéza

- Co je hypotéza?
- Formální vlastnosti hypotézy?
- Lingvistické hypotézy...

„Hypotéza“ v lingvistice

- „V hláskosloví ani v jiných rovinách vodňanského herbáře nejsou prokazatelné další nářeční jevy z oblasti, kde nedošlo ke vzniku vibranty ř (východomoravské území), dáváme proto přednost **hypotéze**, že se jedná o nedbalý zápis“ (Černá 2005, s. 76);
- „V nich se překlad Františka Vrby jeví jako silně zatížený mužským genderovým úhlem pohledu a estetikou vnímání; potvrzuje se tak původní **hypotéza**, že se spíše „staví na stranu“ mužského hrdiny, resp. autorského tvůrce a erotické líčení prezentuje spíše z jeho perspektivy...“ (Širokovská 2004, s. 23);
- „Proč není samo slovo *plémě* ve staročeských textech doloženo v očekávaném významu, o tom lze vznášet různé **hypotézy**.“ (Šimandl 2007, s. 238);
- „**Hypotéza 2.1:** Co-text je věrným zrcadlem (situačního) kontextu v tom smyslu, že všechny pro danou komunikační situaci relevantní kontextové vlastnosti jsou co-textem explicitně reflektovány, a mají tedy nějaký textový korelát. (...) **Hypotéza 2.2:** (Textový) kontext věrně reflektuje všechny vlastnosti jazykových jevů relevantní pro jejich užití. (Cvrček 2013, s. 24)“;
- „Vycházeje z toho, že teorie valence i přes zjevná slabá místa představuje dobrý konstrukt lingvistické teorie, pokusím se nyní představit **hypotézu modifikované valenční teorie** (MVT) a formulovat základní principy této teorie.“ (Karlík 2001, s. 171n).

Od popisu k hypotéze

- předpokládaný vztah mezi dvěma vlastnostmi = působení mechanismu
- zaměříme-li na vztahy mezi vlastnostmi, překračujeme hranice popisu jednotek směrem k odhalování toho, proč jsou sledované vlastnosti oněch jednotek takové, jaké jsou
- nástrojem analýzy tohoto typu je empiricky testovatelná hypotéza

Hypotéza

- hypotéza (Greis 2009, s. 11)
 - tvrzení, které se týká více než jednoho jevu či případu;
 - má alespoň implicitně strukturu podmínkového souvětí, tj. „*jestliže..., pak...*“, případně „*čím..., tím...*“ (např. čím je slovo frekventovanější, tím je kratší);
 - je falzifikovatelné (tj. vyvratitelné) prostřednictvím experimentu, který dovoluje rozhodnout, zda predikce formulovaná prostřednictvím hypotézy je vyvrácena, či ne (vyhodnocení se většinou experimentu pomocí statistických testů).

Hypotéza

- která tvrzení *jsou/nejsou* testovatelnými hypotézami?
 1. *hodně mužů má pleš*
 2. *pokud se v knize vyskytují biblický příběh, je to apokryf*
 3. *jestli se zavedou řidičáky „na zkoušku“, může se snížit nehodovost mladých řidičů a řidiček*
 4. *muži mají častěji pleš než ženy*
 5. *jestliže se je sloveso dokonavé, častěji se na něj váže přímý akuzativní předmět než na sloveso nedokonavé*
 6. *ženy jsou citlivé*
 7. *čím je slovo frekventovanější, tím je větší jeho polysémie*
 8. *jestli se zavedou řidičáky „na zkoušku“, sníží se nehodovost mladých řidičů a řidiček*
 9. *nářečí často ovlivňují podobu mluveného jazyka obyvatel dané nářeční oblasti*

Hypotéza

- pokuste se formulovat lingvistické hypotézy

Hypotéza

- Wikipedie
 - <https://cs.wikipedia.org/wiki/Hypot%C3%A9za>

Hypotéza - opakování

- která tvrzení *jsou/nejsou* testovatelnými hypotézami?
 1. *delší klauze (měřeno v počtu slov) mají v průměru kratší slova (měřeno v počtu slabik) než klauze kratší*
 2. *v odborných textech je hodně dlouhých vět*
 3. *pokud je slovo syntakticky závislé na substantivu, je to přívlastek*
 4. *auxiliáry jsou v průměru kratší než autosémantika*
 5. *mezi délkou slova měřenou v počtu hlásek a v počtu slabik je lineární závislost*
 6. *děti z měst mají bohatou slovní zásobu*
 7. *čeština je jeden z nejkomplicovanějších jazyků na světě*
 8. *čím je slovo delší, tím má více hlásek*
 9. *čím je člověk starší, tím v průměru používá více zájmen*