



# Srozumitelnost textu Výzkum a praxe

Silvie Cinková

[cinkova@ufal.mff.cuni.cz](mailto:cinkova@ufal.mff.cuni.cz)



Program vzdělávání Srozumitelný úřad pro úředníky územních  
samosprávných celků, Ministerstvo vnitra ČR, 23. 9. 2021  
S podporou GAČR 19-191915 „Lingvistické aspekty srozumitelnosti českých právních a úředních textů“

Díky za pozvání a díky, že jste se do toho pustili!  
Silvie Cinková: MFF – NLP, strojové učení, data  
pro strojové učení („umělá inteligence“). Váš  
projekt nás vědecky zajímá & chceme být  
užiteční

Výzk. proj. s J. Šam a Cam Lat. – gačr

Srozumitelnost je fascinující téma – dovoluji mi ho  
představit v širších souvislostech, možná vás  
pak i ten kurz bude víc bavit

## Přehled

- Výhody srozumitelnosti: úspory, důvěryhodnost, fungující společnost
- Výzkum srozumitelnosti
- Umělá inteligence ve službách efektivní úřední komunikace



Zdeněk Peška, profesor práva, 1939 Jazyk zákonodárcův ve Slově a slovesnosti:

... zdálo by se [...], že jazyk může splnit funkci jazyka právního jen tehdy, když je oběma stranám stejně srozumitelný [...] Avšak takové závěry by byly chybné. [...] V moderním státě totiž zákonodárce zpravidla nepočítá s tím, že by text zákona čtlo veškeré občanstvo. Text samého zákona je znám pouze určitému okruhu osob, t.j. obyčejně specialistům [...] a určité části úřednictva, pověřeného jeho prováděním [...]

Pro Peška bylo normální, že zákony byly psány německy a maďarsky (v cizím jazyce!).

Anglosaská právní tradice dokazuje, že se Peška

## Známé „triky“ pro srozumitelné psaní

- spíše kratší věty
- text sekat do odstavců
- používat nadpisy
- mezi větami spojovací výrazy (*tedy, ale, proto...*)
- slovesa raději než podstatná jména (*předkládáme návrh -> navrhuje*)
- činný rod raději než trpný
- výčty raději jako seznamy s odrážkami
- obracet se přímo na čtenáře (druhá osoba, oslovení)



Tohle ale nebude stačit. Ukážeme si.

## Úspory času a práce + přesvědčivost

Před	Po
§ 95.455 Povolené frekvence	§ 95.407 Jaké kanály smím využívat?
§ 95.457 Politika zpřístupňování frekvencí	§ 95.408 Jak vysoko si smím dát anténu?
§ 95.437 Omezení anténových struktur	§ 95.409 Jaké vybavení smím použít?
§ 95.511 Údržba a servis přenosových zařízení § 95.613 Spotřeba přenosových zařízení	§ 95.410 Kolik proudu smím spotřebovat?
§ 95.509 Zákaz externích zesilovačů rádiových frekvencí	§ 95.411 Smím používat zesilovače?

- Joseph Kimble: *Writing for Dollars, Writing to Please*. Scribes Journal of Legal Writing 6/1. 1996.
- G. Mills & M. Duckworth: *The Gains from Clarity*. A research report on the effects of plain-language documents. Centre for Microeconomic Policy Analysis and Centre for Plain Legal Language. Sydney. 1996.
- Anneli Hedlund: *Klarspråket lönar sig*. Stockholm, 2006



Tabulka: víc než „překlad“, je to digest/výtah z právních předpisů orientovaný na potřeby čtenáře – už „předžvýkáno“, co z předpisů vyplývá prakticky. Dedukci od obecného ke konkrétnímu už provedl osvědčený úředník.

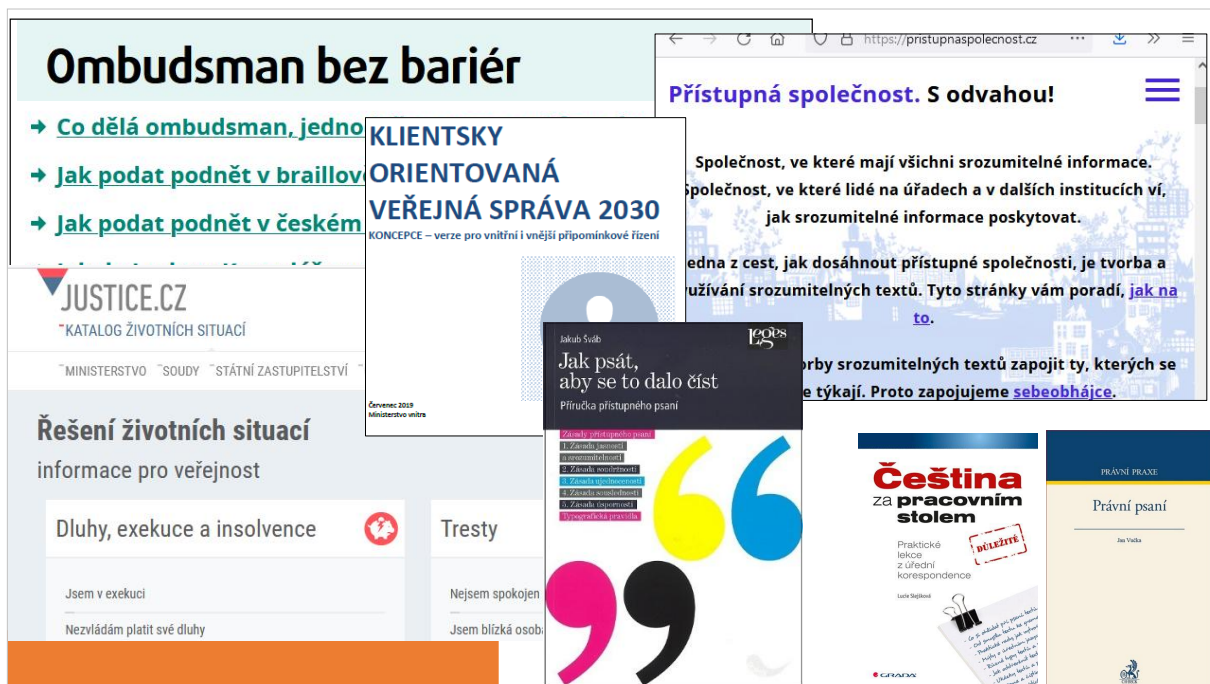
Výhody takového postupu se dají vyjádřit v penězích. Příklady pocházejí z těchto tří publikací – existuje jich ale mnohem víc. První dvě položky: právníci, Arthur Levitt byl předsedou úřadu U.S. Securities and Exchange Commission (info o společnostech obchodovaných na burze).

- Austrálie: rozvodový formulář – sami vyplnili před zjednodušením 52%, po 67%; odmítnutých formulářů z 30% na 7%. Mills & Duckworth
- USA:

## „Plain Language“ ve světě

- USA: 2010 Plain Language Act – úřady musí zajistit srozumitelné dokumenty u všech nařízení práv a povinností občanů, zejména sociální podpory a daní (<https://www.plainlanguage.gov/law/>)
- Velká Británie: úřady mají svá oddělení pro srozumitelnost
- Skandinávské země, Estonsko: „Jazykový úřad“/ombudsman v agendě i srozumitelnost, (standardy, osvěta, kontrola kvality, online korektory, asistence občanům)
- Německo: IDEMA – Internet-Dienst für eine moderne Amtssprache
  - správní celky sdílejí databázi zdařilých dokumentů, slovníky, odbornou podporu
- Evropská Komise: direktivy/příručky o srozumitelném psaní

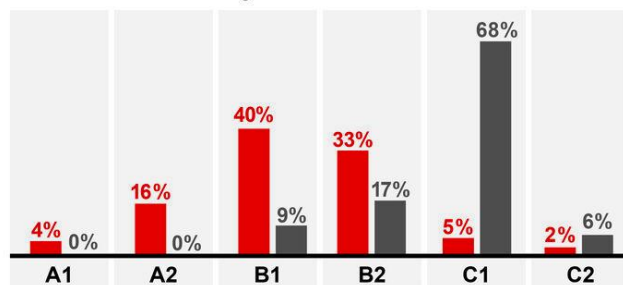
Švédsko má politiku srozumitelného jazyka od 1967, od roku 76 zapojuje jazykovědce, od roku 78 je jazykový poradce pětisemestrální studijní obor



V ČR: situace je nadějná!  
 Pojdme se ale podívat na ten výzkum. Napřed  
 výzkum čtenáře, až potom textu.

## Osoba průměrného rozumu podle CEFR

■ Kompetenzniveau der deutschen Bevölkerung  
■ Sprachniveau der Behörden und Firmeninformation



Quelle: R. Beekveldt, Spaß am Lesen Verlag

převzato z <https://www.capito.eu/richtige-sprachstufe-finden-und-texte-besser-formulieren/>

### Úroveň čtení B1:

Rozumím textům, které obsahují slovní zásobu často užívanou v každodenním životě nebo které se vztahují k mé práci. Rozumím popisům událostí, pocitů a přáním v osobních dopisech.

<https://europa.eu/europass/system/files/2020-05/CEFR%20self-assessment%20grid%20CS.pdf>



Obč. zákoník mluví o osobě průměrného rozumu jako konzumentovi/objektu práva. Kdo je osoba průměrného rozumu?

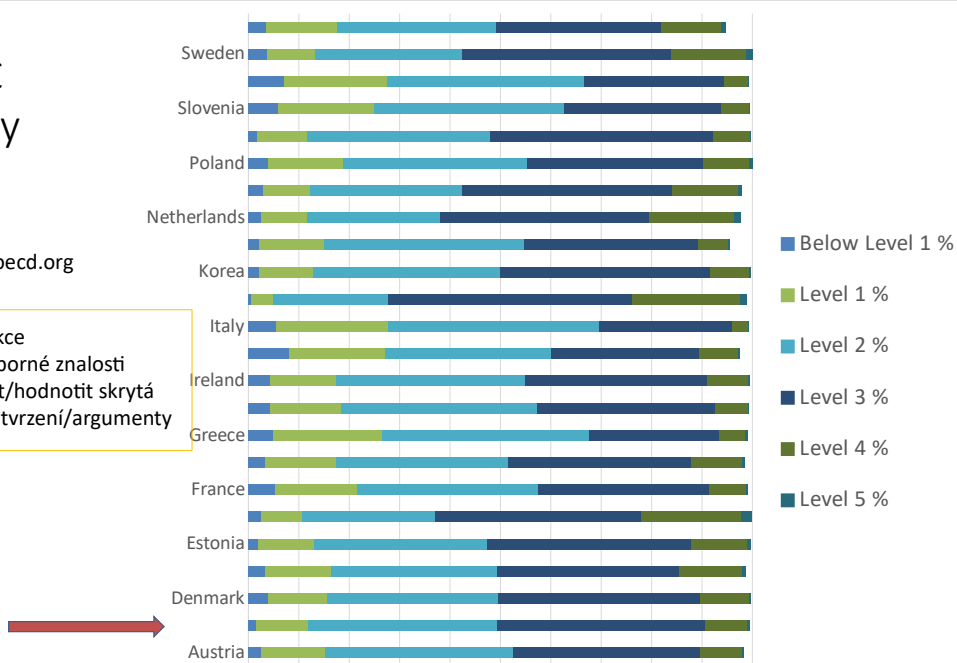
Přiblížíme si na dvou mírách: jazyková kompetence a čtenářská gramotnost

## OECD PIAAC Adult literacy 2019

<https://gpseducation.oecd.org>

### 4

- složité dedukce
- aplikovat odborné znalosti
- interpretovat/hodnotit skrytá pravdivostní tvrzení/argumenty



## Programme for the International Assessment of Adult Competencies

20 úloh,

Čtenářská gramotnost: ČR – průměr EU, velmi malý rozptyl, 5. percentil 1.-3. nejvyšší – i špatní jsou relativně dobří

Úroveň 3: webový knihovní katalog, určit jméno autora knihy Ecomyth, musí prolistovat několik stránek, rozvržení stránky je trochu nepřehledné.

**Úroveň 4:** aby našel knihu, která naznačuje, že argumenty pro i proti geneticky modifikovaným potravinám jsou nespolehlivé. Přečíst název a popis všech knih. U jedné se praví, že autor „popisuje, jak obě strany ostře diskutované otázky vytvářejí propagandu, snaží se oklamat veřejnost a ... [zde text končí]“.



https://www.oecd.org/skills/ESonline-assessment/takethetest/

The screenshot shows the top navigation bar of the OECD Education & Skills Online Assessment website. On the left, there are the OECD logo (with the tagline 'BETTER POLICIES FOR BETTER LIVES') and the European Commission logo. On the right, there are social media icons for email, Twitter, Facebook, and YouTube, along with a language dropdown menu set to 'English' and a search bar. Below the navigation bar is a blue banner with the text 'Education & Skills Online Assessment' and a globe background. At the bottom of the banner is a horizontal menu with five items: 'ABOUT EDUCATION & SKILLS ONLINE', 'SKILLS ASSESSED', 'ASSESSMENT DESIGN', 'ASSESSMENT ADMINISTRATION', and 'TAKE THE TEST'.

Please click on the language of your choice to begin Education & Skills Online.

Use our DEMO version to see sample test questions and answers

**You must use Firefox – most current version or one version older than most current.** The test does not function properly in other browsers.

To take the test, you must have received an authorisation code.



Test in Czech (Ceřtina)

[Demo Test](#)



Test in English (Australia)

[Demo Test](#)



Test in English(Canada)

[Demo Test](#)

## If it's easy to read, it is easy to do, pretty, good, and true (Song & Schwarz 2010)

- Čtenář nerozliší obtížnost čtení (i typografie!) od obtížnosti obsahu.
- Špatná přístupnost textu vede k neochotě provést instrukci, odkládání rozhodnutí,
- Působí jako příliš obtížné nebo časově náročné, nebo nevěrohodné
- („Fluency & familiarity affect risk and truth assessment.“)



Běžný úřední text tedy většinu čtenářů přetěžuje. Zdroje frustrace: laik nemůže umět spolehlivě identifikovat všechny relevantní informace, souvislosti mezi nimi jsou implicitní – ovlivní to jeho přístup k obsahu. Experimenty z kognitivní psychologie: Song&Schwarz, 2010, The psychologist; M. Oppenheimer The secret life of fluency.

Fluency: subjektivní hodnocení obtížnosti nějakého mentálního úkonu. Evolučně dané: když se informace o něčem špatně zpracovává, bude to obtížné – zvíře špatně viditelné v mlze se loví hůř, než když je vidět. Možná radši půjdu na kořínky. Když je špatně slyšet, je i špatně rozumět obsahu. Přestanu vnímat, někde si to přečtu nebo

# Procedurální spravedlnost

## Návrh zjednodušeného poučení obviněného před výslechem

### ? Výslech

Jakub Drápal:  
Procedurální  
spravedlnost a trestní  
řízení: zahraniční  
poznatky a jejich možné  
využití v ČR.  
Advokátní deník,  
8. 4. 2020

Tom R. Tyler. 1990.  
Why People Obey the Law.  
Yale University Press.

**Máte právo vypovídat** § 33/1 trestního řádu (TRŘ)  
Máte právo, ne povinnost, se ke všemu vyjádřit.  
Můžete tedy odmítnout vypovídat.

**Podrobné souvislé vyjádření** § 33/1, § 92/2 TRŘ  
Máte možnost se podrobně vyjádřit k předmětu  
obvinění.

**Kladení otázek** § 92/3 TRŘ  
Mohou Vám být kladeny doplňující otázky.

**Možnost radit se o odpovědích** § 33/1 TRŘ  
Během výslechu se nemůžete s obhájcem radit  
o tom, jak odpovědět na již položenou otázku.  
Před jejím položením se radit můžete.

**Zákaz klamavých a návodných otázek** § 92/3 TRŘ  
Nesmí Vám být kladeny otázky, při kterých je  
tvrzeno něco, co není pravda. Zároveň Vám nesmí  
být naznačeno, jak máte odpovědět.

### Obhájce

**Přítomnost obhájce** § 33/1 TRŘ  
Máte právo, aby byl Váš obhájce přítomen  
u výslechu nebo jiných úkonů.

**Bezplatný obhájce** § 33/2 TRŘ  
Pokud doložíte, že nemáte dostatek finančních  
prostředků na obhájce, máte právo na přidělení  
obhájce bezplatně, či za sníženou cenu.

### 🔗 Když Vás přivedla policie

**Vyrozumění třetí osoby** § 33/5 TRŘ  
Máte právo dát o svém zadržení někomu vědět.

**Lhůta pro zadržení či zatčení** § 33/5 TRŘ  
Do 24 hodin od zatčení či 48 hodin od zadržení  
musíte být buď předáni soudu, nebo propuštěni.



Když se čtení daří, čtenářovi připadá, že má i  
popisovanou situaci pod kontrolou – uvolní se a  
spolupracuje

## Paměť a zpracování textu (Givón, 1993)

- Pracovní paměť: 8-20 sekund textu / 2-5 klauzí (jednoduchých vět)
- Epizodická paměť, pojímá reprezentaci obsahů z pracovní paměti
- Reprezentace v podobě sítě s uzly („témata“) a hranami („vztahy“)
  - sekvenční (témata jdou za sebou v čase)
  - hierarchická (nová informace se vztahuje k už známému tématu)
  - dynamická (uzly se převěšují, hrany vznikají a zanikají)
- Čím propojenější uzel, tím lépe se v paměti udrží
- Pomáhá i souvislost mimo daný text, emoce...
- Dlouhodobá paměť: sem jde definitivní obraz sítě



Psychologické modely zpracování textu:  
návody na srozumitelné psaní si někdy protiřečí  
(pište to kratší – ne, uvádějte příklady, ty to  
ovšem prodlouží)

– zde je příčina: na zpracování textu má vliv  
mnoho proměnných. Váš cíl: dostat do  
čtenářovy dlouhodobé paměti všechnu  
informaci, kterou má použít k tomu, co po něm  
chcete.

Je to věc cviku se zpětnou vazbou od čtenářů.  
Anekdota s obecním blbem v Rakousku-  
Uhersku.

# Hamburský koncept srozumitelnosti



- Jednoduchost
  - slovní zásoba
  - větná skladba
- Členění uspořádání
  - odstavce, nadpisy, seznamy
  - typografie
  - logika, relevance, „červená nit“
- Délka/hutnost
- Dynamičnost
  - udržování pozornosti
  - příklady

Optimum	
Jednoduchost ++	Hutnost mezi 0 a +
Členění/uspořádání ++	Dynamičnost mezi 0 a +

Potvrzení toho, co píše Givón. Německo, 70. léta, dodnes populární kurzy a učebnice psaní. Hromada parafrázovaných textů, hodnotili jejich vlastnosti na škále -2 +2, pak měřili porozumění na velkém množství respondentů. Jaké vlastnosti textu korelují s porozuměním? Původně hodnotili 40 vlastností, statisticky redukovali na 4, ověřili v dalším kole testů.

# Projekt LiFR

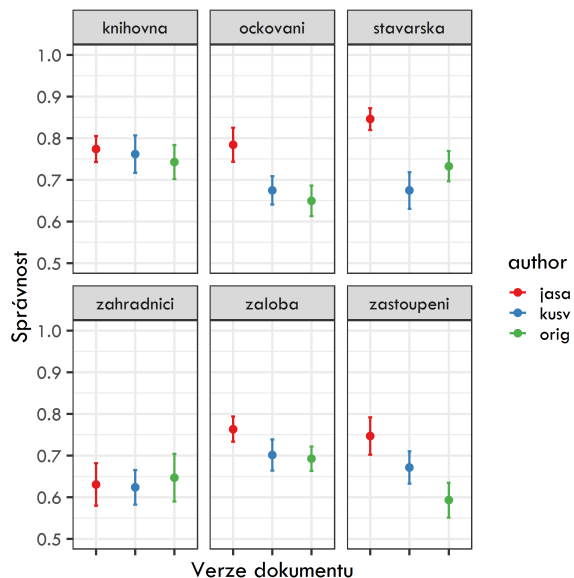
- Šest úředních **dokumentů** ve třech **verzích** = 18 zkušebních **textů**

- 2x ombudsmanská zpráva,
- knihovní řád,
- rozhodnutí soudu o zastupování členem domácnosti,
- žaloba o náhradě škody,
- smlouva o pronájmu

- Každá trojice dokumentů sdílela dotazník porozumění

- test 7 otázek
- subjektivní hodnocení 5 otázek
- otázky na předchozí znalost + demografie

- Test porozumění: 200 respondentů, 800 vyplněných dotazníků



LINDAT

UFA



subjektivní hodnocení zhruba odpovídalo 4 vlastnostem hamburského konceptu (rozdělili jsme slovní zásobu a syntax)  
Lineární smíšený model ukazuje signifikantní a silný efekt autora – tj. autoři píšou různě srozumitelně, napříč téměř všemi texty. Rozdíl není jen u smluv a pravidel.

## Subjektivní hodnocení vs. úspěšnost odpovědí

- Hutnost vyjádření
- Syntaktická složitost
- Přiměřenost slovní zásoby
- Logická koherence
- Snaha autora zaujmout
- Silný efekt subjektivního hodnocení na správnost odpovědí
- Nejsilnější: koherence a slovní zásoba



Opět smíšený lineární model.

# Syntaktická složitost a porozumění

- **Žádná korelace průměrného porozumění s:**
  - délka a složitost klauzí
  - nominalizace
  - postponované a participiální přívlastky
  - počet souřadných výčtů
  - metriky srozumitelnosti (Flesch Reading Ease, Flesch Kincaid, ARI, Coleman-Liau)
- Ani subjektivně pocíťovaná složitost neměla velký vliv!

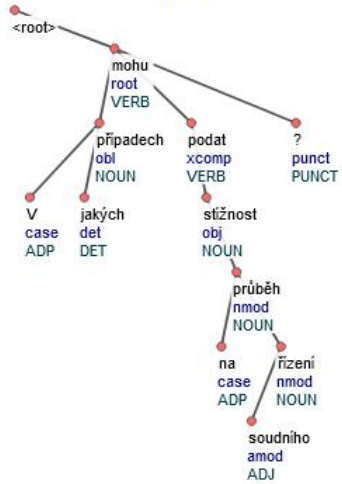




# UDPipe

[About](#) [Run](#) [REST API Documentation](#)

V jakých případech **mohu** podat stížnost na průběh soudního řízení ?



# Umělá inteligence radí... s němčinou

- capito app (<https://www.capito.eu/>)

Voraussetzungen für eine nachträgliche ordentliche Veranlagung:

Sie verfügen über weitere nicht der Quellensteuer unterliegende Einkünfte über CHF 3'000 bzw. Vermögen über CHF 80'000 (bei gemeinsam steuerpflichtigen Personen CHF 160'000). Wir weisen darauf hin dass Sie in diesem Fall verpflichtet sind einen Antrag auf nachträgliche ordentliche Veranlagung zu stellen.

Sie möchten eine freiwillige nachträgliche ordentliche Veranlagung vornehmen um abzugsfähige Aufwendungen geltend zu machen, welche in den Quellensteuertarifen nicht oder nur pauschal eingerechnet sind.

Dies können beispielsweise folgende Abzüge sein:

- effektive Berufskosten (inkl. Wochenaufenthaltskosten)
- Einkäufe in die berufliche Vorsorge (Pensionskasse BVG / 2. Säule)
- Beiträge an die gebundene Selbstvorsorge (Säule 3a)
- Selbstgetragene Aus- und Weiterbildungskosten bis maximal CHF

capito digital

192 Wörter

The screenshot displays the Capito app interface. On the left, a German tax document is visible with highlighted sections. The app's AI analysis tools are shown on the right, including a 'Test-Abo' notification, a list of detected entities (Voraussetzungen, Sie, verfügen, Quellensteuer, unterliegende, CHF), and a sidebar with user profile information (Silvie Cinkova) and language settings (Sprachstufe wählen: B1, A2, A1). The app also features filters for simplifications (Wortwahl, Layout & Text, Gendern Beta, Hinweis) and a 'Jetzt anfragen' button.

LINDAT

UFAL



Capito AI ji zřejmě může modelovat z dat za 20 let  
„nechte nás přeložit“  
„začněte psát a nechte nás to opravit“

## Hypotéza: **všemu dominuje efekt koherence**

- Výběr textů pro LIFR: nenašel se koherentní syntakticky kostrbatý text.
- AI: modelovat argumentační vzorce koherentních textů





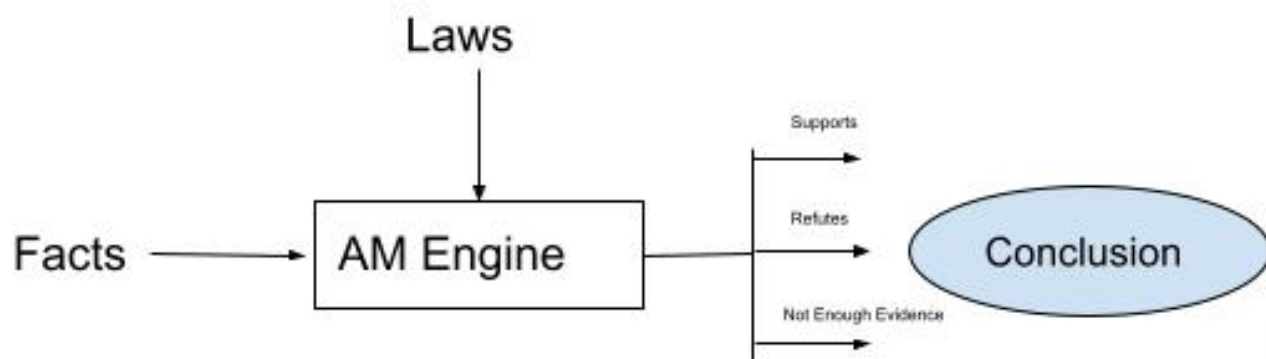
## Modelování argumentačních vzorců: Sylogismus

- Obecný předpoklad
  - Konkrétní předpoklad
  - Závěr
  - Právní pravidlo
  - Skutkový stav
  - Právní hodnocení
- 
- Všichni lidé jsou smrtelní.
  - Sokrates je člověk.
  - Proto je Sokrates smrtelný.
  - Matkou dítěte je žena, která je porodila (§ 775 OZ).
  - Marie Vomáčková dítě porodila.
  - Proto je Marie Vomáčková matkou dítěte.

Důležité k tomu:

# Argumentation Mining

Teruel et al., (2018) Korpus rozhodnutí  
Evr. soudu pro lidská práva – anotace  
předpokladů, tvrzení a jejich vztahů  
(podporuje - napadá).



# Data pro češtinu?

- paralelní korpusy textů „před“ zjednodušením a „po“ zjednodušení
- neparalelní korpusy „srozumitelných“ a „nesrozumitelných“ textů
- klasifikace podle typů, např.
  - kvaziobjektivní (soudní/úřední rozhodnutí)
  - persuasivní (žaloba, podání soudu)
  - normativní (vyhlášky, zákony, podobné i smlouvy)

## Možné aplikace

- Interaktivní pomocník pro právní psaní
- Automatická sumarizace nebo vytváření otázek k textu
- Extrakce informací
  - do doménových ontologií
  - vyhledávání argumentačně/obsahově podobných dokumentů (vzory dokumentů)



Děkuji za pozornost.  
cinkova@ufal.mff.cuni.cz

