Základy využití korpusu v praxi **cjbb75**

**Středa: 9.10-10.50 G13**

**3. 5. 2017**

**Opakovací hodina – podobné úkoly budou na písemce. Na hodině se patrně nestihne udělat vše, takže za dú si zbytek doděláte. Vzhledem k tomu, že 10. 5. 2017 je dopoledne rektorské volno, nemusíte úkoly odevzdávat. V IS poskytnu správné řešení – zpětnou vazbu.**

1. Končí-li tvar slova v češtině na ***–ové***, pak jde nutně o tvar životného maskulina. Platí toto tvrzení? Odpovězte nejdříve, a pak ověřte svoji odpověď v korpusu SYN2015. Součástí odpovědi musí být formulace dotazu, pomocí kterého byla získána data, na jejichž základě odpovídáte.
2. Dokažte, že není třeba, abyste věděli, co označuje vyžlucené substantivum (abyste četli H. Pottera) k tomu, abyste byli schopni na základě zadání samého vyřešit následující úkol. **Na základě uvedených vět určete rod a vzor vyžluceného substantiva.**

Mudlové (v originále Muggles) je označení pro lidi, kteří neumí kouzlit a jsou nekouzelnické nebo smíšení krve. Mudlaje tu pro vás.

1. Hláska *e* má v češtině dvě grafické realizace: *e/ě*. Existuje český slovesný tvar, který končí na grafickou realizaci *ě*? Odpovězte nejdříve, a pak ověřte svoji odpověď v korpusu SYN2015. Součástí odpovědi musí být formulace dotazu, pomocí kterého byla získána data, na jejichž základě odpovídáte.
2. *Místo vašich schůzek sis měl lépe vybírat.* Kolik interpretací má slovní tvar *místo* v uvedené větě? Podívejte se do korpusu SYN2000 a zjistěte, jak jsou desambiguovány interpretace slovního tvaru místo. V případě, že naleznete chybné interpretace, pokuste se navrhnout postup, kterým byste je detekovali.
3. Názvy zlomků (substantiva tvořená ze základů číslovek základních sufixem *–ina*, např. *pětina, desetina, dvanáctina, padesátina, …*) a názvy jubileí (*padesátiny, třicetiny, šestnáctiny, …*) mají část tvarů společných (plurálové tvary), což způsobuje nepříjemnosti, chceme-li mít oba významy odlišeny na rovině automatické morfologické analýzy. V korpusu SYN2015 vyhledejte relevantní lemmata a pozorujte chyby v desambiguaci. Navrhněte postup, jak nalézt případy, které by bylo možné zjednoznačnit na základě blízkého kontextu.
4. V české flexi platí pravidlo, které tvrdí, že substantiva skloňovaná podle vzoru *růže* mají v genitivu plurálu buď koncovku *–í* (*růží*), nebo koncovku *–0* (*vesnic*). Dokázali byste odpověď na následující otázku, kterou by vám položil někdo, kdo se chce naučit česky (cizinec) a která by zněla: Znamená to, že se u každého českého slova, které je ženského rodu a končí v nominativu sg. na *–e* musím naučit, jak vypadá jeho tvar v genitivu plurálu, nebo mi můžete ještě pomoci nějakým dalším rozšířením pravidla uvedeného výše? Odpovězte nejdříve, a pak ověřte svoji odpověď v korpusu

SYN2015. Součástí odpovědi musí být formulace dotazu, pomocí kterého byla získána data, na jejichž základě odpovídáte.

1. Dokázali byste odpověď na následující otázku, kterou by vám položil někdo, kdo se chce naučit česky (cizinec) a která by zněla: Učil jsem se, jak se tvoří tvar l-ového příčestí od sloves, který musím znát, abych uměl vytvořit větu v minulém čase a v podmiňovacím způsobu. Všiml jsem si, že někdo říká, že *počítač vypl* a někdo, že *počítač vypnul.* Jak poznám, když se učím nové sloveso, které má v infinitivu zakončení na *–nout*, jak má vypadat tvar příčestí? Odpovězte nejdříve, a pak ověřte svoji odpověď v korpusu SYN2015. Součástí odpovědi musí být formulace dotazu, pomocí kterého byla získána data, na jejichž základě odpovídáte.
2. V rámci publikací ÚČNK vyšla i tato publikace: Cvrček, V., Cvrčková Porkertová, L.: Velký slovník rýmů. Nakladatelství Lidové noviny, Praha 2011. Její anotace zní: Příručka rýmů usnadní psaní jakýchkoli rýmovaných textů, což je úkol, před kterým stojí nejen začínající básníci, ale také autoři písňových textů, reklamních sloganů a mnozí další. Slovník zároveň může sloužit jako doplňková pomůcka při výuce literární teorie na základních a středních školách. Svým rozsahem i způsobem zpracování se jedná o unikátní projekt, který na současném knižním trhu nemá obdoby (posledním takovým pokusem byl Puchmajerův Rýmovník z roku 1824).

Nemáte slovník po ruce. Zaexperimentujte se svojí jazykovou vynalézavostí a s korpusem. Najděte (po paměti) rýmy slov *běhna*, *loutna, kout*. Vyhledejte v korpusu synv4 doklady, které mohou být rýmem slova/tvaru a porovnejte svoji jazykovou vynalézavost s tím, co lze najít v korpusu. Použili jste jako pomůcku k vyhledávání v paměti nějaký rýmovaný text, který umíte zpaměti? Pokud ano, uvědomte si, nakolik byl autor textu v otázce rýmu vynalézavý (rýmové slovo má v korpusu velkou – malou – žádnou frekvenci).

1. Tabulkové paradigma tvarů *by*, které v češtině slouží ke tvoření kondicionálu bývá v českých mluvnicích uvedeno následovně:

|  |  |  |  |
| --- | --- | --- | --- |
| osoba | singulár | osoba | plurál |
| 1. | ***bych*** | 1. | ***bychom*** |
| 2. | ***bys*** | 2. | ***byste*** |
| 3. | ***by*** | 3 | ***by*** |

Bez komentáře k tabulce, by tudíž následující tvrzení: „**Tvar *by* může být v češtině součástí tvaru kondicionálu 2. osoby singuláru**“ nemělo platit. Platí, nebo neplatí? Pokuste se odpovědět na základě introspekce. Pokuste si ověřit svoji odpověď v korpusu SYN2015.

1. I když je letec **více DiCapriův než Scorseseho** film, je zřejmé, že pro Scorseseho, vždy fascinovaného filmem i zuřivými hrdiny, to byla mnohem více než pouhá zakázka.

Podívejte se na větu z korpusu synv4 Pokuste se najít podobné doklady a zamyslete se nad tím, zda jde o analytické stupňování subjektově posesivních adjektiv.