

Abstrakt

V práci (Hlaváčová 2009) je uvedena množina šesti českých slovesných předpon, které spolu s reflexivním morfémem způsobí modifikaci původního slovesa. Nemodifikuje se však hlavní význam slovesa, ale intenzita děje, který je slovesem vyjádřen. Jedná se o předpony roz-, po-, za-, na-, vy- a u- s reflexivním morfémem se nebo si (u předpon po- a za-). Uvedené pořadí současně zavádí uspořádání předpon podle intenzity slovesného děje. Předpona roz- tedy znamená začátek, malou intenzitu, předpona u- potom intenzitu nejvyšší. Intenzity vyjádřené pomocí ostatních předpon se mohou překrývat, jejich stupeň není ostře vyhraněný.

Tento typ slovesné modifikace je velmi produktivní, lze ho aplikovat na většinu nedokonavých sloves. V dalších pracích (Hlaváčová, Nedolužko 2012, 2013) se ukázalo, že podobné vlastnosti mají tyto předpony i ve slovenštině a v ruštině; z nepublikovaných konzultací se potom lze domnívat, že i polština a chorvatština má podobnou posloupnost intenzifikujících slovesných předpon. Nabízí se tedy hypotéza, že uvedený způsob modifikace slovesa je vlastností více (všech?) slovanských jazyků. V našem příspěvku se však chceme zabývat jen češtinou.

V uvedených pracích je předloženo ke každé předponě několik příkladů - vymyšlených i převzatých z korpusů či z internetu. Není tam však předvedeno, zda je možné uvedené typy sloves v běžném textu správně rozpoznat. Některá slovesa s příslušnou předponou a výše popsáním významem jsou běžnou součástí jazyka, v jiných případech však jde o slova příležitostná, která přirozeně nejsou zahrnuta v žádném slovníku. Jedním z důsledků je to, že zůstávají v jazykových korpusech většinou nerozpoznána. Při automatickém zpracování jazyka je však výhodné i takové slovní formy umět rozpoznat a přiřadit jim správné morfologické vlastnosti. Rozšířit slovníky o všechny možné tvary takových sloves však není příliš užitečné, vezmeme-li v úvahu jejich poměrně nízkou frekvenci.

V tomto příspěvku bychom tedy chtěli ukázat, jak taková slovesa v textech vyhledávat, a stanovit kritéria, která umožní jejich správné určení. Pomocí klasických měř pro vyjádření přesnosti ukážeme, jak jsou takové vyhledávky úspěšné. Vše bude prezentováno na datech Českého národního korpusu.

Konkrétně jsme tedy navrhli vyhledávací podmínky, které v anotovaném korpusu řady SYN označí všechna slova, jejichž zápis začíná jednou z uvedených předpon a v jejichž rozumném okolí se nachází reflexivní morfém se (případně si). Ve výsledné množině se samozřejmě objeví i jiná slova, včetně překlepů, kterými se zabývat nechceme. Proto výslednou množinu dále zúžíme. Odtrhneme předponu a spustíme morfologickou analýzu na zbylý řetězec. Vycházíme z předpokladu, že základní sloveso, ke kterému je připojena předpona, je běžnou součástí jazyka, a tudíž obsaženo v morfologickém slovníku. Jestliže tedy analýza určí, že zbytek je nedokonavé sloveso, zařadíme tento výskyt do výsledné množiny. Ukazuje se, že tyto množiny nejsou příliš rozsáhlé, obsahují maximálně řádově desítky kandidátů, takže je snadné je ručně probrat a rozhodnout, zda se jedná o sloveso s daným významem, či nikoli. Z výsledných počtů jsme potom pro všechny předpony spočítali hodnotu tzv. precision, která určí, jaký poměr z vybraných kandidátů je skutečně hledaným slovesem.

Výsledky ukázaly, že některé předpony mají hodnotu precision poměrně vysokou, což znamená, že jejich vyhledání je spolehlivé. U sloves s předponami, jejichž hodnota precision je nízká, není jejich intenzifikační význam tak jednoduše rozpoznatelný. U některých předpon bude možné využít i jiná kritéria, vycházející především z kontextu. Objeví-li se například v okolí reflexivního slovesa s předponou u- výraz "k smrti", bude jeho určení zřejmě mnohem jistější (např. uhonil se k smrti).

Bibliografie

Hlaváčová Jaroslava: Formalizace systému české morfolgie s ohledem na automatické zpracování českých textů. Ph.D. thesis, FF UK, Nám. Jana Palacha 2, 116 38 Praha 1, 146 pp., May 2009

Hlaváčová Jaroslava, Nedolužko Anna: Příklad pravidelných slovotvorných vzorců v automatickém zpracování češtiny a ruštiny. In: Zborník príspevkov prezentovaných na konferencii Informačné technológie – Aplikácie a Teória, ITAT 2012, Hotel Magura, 17-21. septembra 2012, Copyright © Slovenská spoločnosť pre umelú inteligenciu, Košice, Slovakia, ISBN 978-80-971144-1-1, pp. 53-56, 2012

Hlaváčová Jaroslava, Nedoluzhko Anna: Intensifying Verb Prefix Patterns in Czech and Russian. In: Lecture Notes in Computer Science, Vol. 8082, Text, Speech and Dialogue: 16th International Conference, TSD 2013. Proceedings, Copyright © Springer Verlag, Berlin / Heidelberg, ISBN 978-3-642-40584-6, ISSN 0302-9743, pp. 303-310, 2013

Korpus SYN2010: Český národní korpus - SYN2010. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z WWW: <<http://www.korpus.cz>>.

Hlaváčová Jaroslava, Nedoluzhko Anna: Productive verb prefixation patterns. The Prague Bulletin of Mathematical Linguistics, No. 101, Univerzita Karlova v Praze, Prague, Czech rep., ISSN 0032-6585, pp. 111-122, 2014

Khoroshkina, Anna and Anna Nedoluzhko. “Vchera nasochinyalsya voroh strok”: Productive circumfixal intensifying patterns in Russian. In Papers from the Annual International Conference Dialogue 2014, volume 13 (20) of Computational Linguistics and Intellectual Technologies, Moscow, 2014.