

HQSYN16 - Task #4006

Task # 3672 (Assigned): RA1d - Automatic cleaning of speech corpora

Task # 3704 (New): Detection and correction of prosodic structures

Task # 4155 (Closed): Prosodic word parsing

Task # 3941 (Closed): Create synthesis logs for corpora with altered prosodic words

Experiments on prosodic words: Synthesis of selected sentences

15.08.2016 16:17 - Hanzlíček Zdeněk

Status:	Closed	Start date:	15.08.2016
Priority:	Normal	Due date:	
Assignee:	Hanzlíček Zdeněk	% Done:	0%
Category:		Estimated time:	0.00 hour
Target version:	RA1: Analysis of artifacts in synthetic speech		
Description			
Please synthesize the following selected sentences for speaker AJ with all 3 variants of prosodic words:			
<p>To je další známka zvýšeného zájmu mezinárodních koncernů o Austrálii. Včera o tom rozhodla vláda na základě doporučení výběrové komise. Zmíněné společnosti patří mezi největší poradenské firmy na světě. Druhy žijící v menších hloubkách prchají v oblaku tmavého inkoustu. Uvedl to politický předák kosovských Albánců Ibrahim Rugova. Pražská Lucerna je druhou zastávkou souboru v rámci letošního turné. Zástupci zemědělců však považují cenu čtyř tisíc za likvidační. Nynější paragrafy už život na vysokých školách dávno předběhl. Podstatné bude přijmout správné rozhodnutí v kauze Matiční. Dva roky dělí francouzské fotbalisty od dvou velkých triumfů.</p> <p>Český obranný průmysl nemůže být izolovaným ostrůvkem. Ženy jsou pak touto chorobou postiženy dvakrát častěji než muži. Bez gólu skončila utkání Plzeň Slavia a Kladno Vítkovice. Cílem služby je odlehčit přetíženému zdravotnímu systému. Desítky let Kanada opomíjela svá rozsáhlá území na severu. Dlouhodobý výhled pro evropské trhy je mírně optimistický. Do schválení nového rozpočtu však ještě zbývá devět měsíců. Dorostu sotva odrostlí hráči ale za Černým pro radu nechodí. Dosáhnout této hranice pro organismus znamená velkou námahu. I střevní problémy mohou mít akutní nebo chronický charakter.</p> <p>Jít do rizikového výzkumu je totiž pro ně drahé. K řešení ovšem musí přispět vláda i vývozci. Konkrétní údaje však ministr odmítl sdělit. Konstantou je však ta vroucnost a opravdovost. Mlčení ale může být pouze zastíracím manévrem. Místní Romové mají však z Albánců stále strach. Míša Řehák protestuje proti těžbě ropy z moře. Na jejím rozumu možná závisí budoucnost Polska. Navíc mu přibýly nečekané potíže se živobytím. Nemocniční lékaři stále volají po vyšších mzdách.</p> <p>Dokladem o této skutečnosti je například zplnomocnění daňového poradce poplatníkem. Okolo tří tisíc vagonů mají železničáři odstaveno na severočeských železnicích. Společnosti nevyplácejí dividendy kvůli svým nedobrym hospodářským výsledkům. Po Kováčově návratu na Slovensko zahájila jeho stíhání bratislavská prokuratura. Bezkonkurenčně nejlepší sněhové podmínky panují ve Francii koncem ledna. Oficiální zpráva tříčlenného obchodního rozhodčího soudu má přijít zítra. Zeman odmítá měsíc odpovídat na otázky týkající se bamberských jednání. Jeho poslední krok se však dotkl zájmů organizované internetové komunity. Atmosféra školního vyučování tak připomínala spíše vojenský režim než hru. Nejčastěji užívanou vyučovací metodou obecných škol bylo hlavně memorování.</p>			

Vláda přitom se svým souhlasem váhala.
K této absurditě v minulosti již došlo.
Proti tomu však nikdo ve vsi neprotestuje.
Vklady u bank jsou totiž zákonem pojištěné.
Ministerstvo čekají též personální změny.
Rychetský je připraven se do věci plně vložit.
Žijeme ve spáleném krámě a ve snědené zemi.
Lhůta pro rozšíření žaloby však již uplynula.
Musel jsem si na živobytí vydělávat tlumočením.
S konečnou platností o prodeji rozhodne vláda.

Věno se vyplácí rodině nevěsty.
Nová verze bude připravena v říjnu.
Boj pojišťoven o motoristy vrcholí.
Domácí kolektiv zachránila přestávka.
O nic jiného také ve skutečnosti nejde.
Tady už dávno běží tendr na jeho kupce.
Avšak rozhodně je stále na čem pracovat.
V rozbitém domě leží ohořelá mrtvola muže.
Nejméně tři kaple budou opraveny již příští rok.
On se tady k tomu asi stejně vyjadřovat nebude.

History

#1 - 16.08.2016 15:01 - Tihelka Dan

- Status changed from New to Assigned

#2 - 05.09.2016 13:07 - Tihelka Dan

- Status changed from Assigned to Resolved

- Assignee changed from Tihelka Dan to Hanzlíček Zdeněk

Files were generated and stored to Experiments/pros_words/out_\$(spkr)_\$(ver) directories (see [#3941#note-2](#) and [#3941#note-3](#)).

#3 - 23.09.2016 12:40 - Hanzlíček Zdeněk

- Status changed from Resolved to Assigned

- Assignee changed from Hanzlíček Zdeněk to Tihelka Dan

Please, repeat the synthesis process for both speakers with the fixed prosodeme bug.

#4 - 14.10.2016 10:42 - Hanzlíček Zdeněk

- File *aj_selection.txt* added

- File *mr_selection.txt* added

Please synthesize new sentences listed in attached files. For each speaker synthesize all sentences (from corresponding file) with all 3 versions of prosodic words.

#5 - 24.10.2016 12:58 - Tihelka Dan

- Status changed from Assigned to Resolved

- Assignee changed from Tihelka Dan to Hanzlíček Zdeněk

Done, outputs are as described in earlier (in [#4006#note-2](#))

#6 - 25.10.2016 16:07 - Hanzlíček Zdeněk

- Status changed from Resolved to Feedback

- Assignee changed from Hanzlíček Zdeněk to Tihelka Dan

Some utterances contain suspicious units in functionally involved prosodemes, e.g. for speaker AJ version 2

- Od Nového roku bude dražší také uhlí. (contains P2)
- Mezi firmami zuří konkurenční boj. (contains P2)
- Stavba má být dokončena zhruba za rok. (contains P0)

- Uvedl to Český statistický úřad. (contains P0)
etc.

Please check whether the selection procedure works correctly.

#7 - 26.10.2016 13:34 - Tihelka Dan

As far as can be seen from the logs, the required prosodeme is 1, but selected unit contains prosodeme 2. Although this mismatch increases target cost, the selection may prefer this unit due to the concatenation cost, which would lead to even higher total cost if another unit is selected. At least it is my guess, we do not have a tool to check it now (neither [PROKUS](#) has this functionality).

#8 - 26.10.2016 17:21 - Tihelka Dan

OK, there really was an error: during TC computation, the costs were not multiplied by their weights, while the total TC was divided by the sum of weights. Therefore, the overall TC was much smaller than it should be and consequently CC was much preferred.

Regenerating the data ...

#9 - 31.10.2016 09:17 - Tihelka Dan

- Assignee changed from Tihelka Dan to Hanzlíček Zdeněk

Could you look at results in out_aj_1, if it is correct now?

#10 - 10.11.2016 14:39 - Tihelka Dan

- Status changed from Feedback to Resolved

For the record: the fixed logs were regenerated for all experiments.

#11 - 15.11.2016 13:54 - Hanzlíček Zdeněk

- File aj_selection_2.txt added

- File mr_selection_2.txt added

- Status changed from Resolved to Assigned

- Assignee changed from Hanzlíček Zdeněk to Tihelka Dan

Please synthesize newly selected sentences listed in attached files.

#12 - 16.11.2016 08:58 - Tihelka Dan

- Status changed from Assigned to Resolved

- Assignee changed from Tihelka Dan to Hanzlíček Zdeněk

Done.
Files were generated and stored to Experiments/pros_words/out_\$(spkr)_\$(ver) directories, as before (see [#3941#note-2](#) and [#3941#note-3](#)).

#13 - 26.01.2017 10:36 - Matoušek Jindřich

- Status changed from Resolved to Closed

Files

File Name	Size	Date	Assignee
aj_selection.txt	3.76 KB	14.10.2016	Hanzlíček Zdeněk
mr_selection.txt	3.49 KB	14.10.2016	Hanzlíček Zdeněk
aj_selection_2.txt	3.99 KB	15.11.2016	Hanzlíček Zdeněk
mr_selection_2.txt	4.03 KB	15.11.2016	Hanzlíček Zdeněk