



MERLIN

**Multiligvální platforma pro evropské referenční
úrovně: mezijazykový výzkum v kontextu**

Uživatelský manuál

Verze 1

11/2014



Projekt MERLIN je financován Evropskou komisí. Webstránky tohoto projektu obsahují pohled na problematiku autorů projektu, přičemž Komise nenese odpovědnost za používání informací v něm obsažených.

Citace projektu: MERLIN project (2014): *User Manual*. www.merlin-platform.eu.

Obsah

Přehled tabulek/grafiky	3
1 Projekt MERLIN – cíl a motivace	5
2 Metodologie práce	7
2.1 Sběr dat.....	7
2.2 Transkripce	7
2.3 Opětovné hodnocení.....	8
2.4 Manuální anotace.....	9
2.4.1 Cílové hypotézy (TH).....	9
2.4.2 Anotace rysů žakovského jazyka	11
2.5 Automatické anotace.....	16
2.6 Kontrola kvality.....	17
3 Sekce dokumentace platformy MERLIN	19
3.1 <<MERLIN: v praxi>>.....	19
3.2 <<MERLIN: výzkum>>	20
3.3 <<MERLIN: korpus>>	22
3.4 <<MERLIN: anotace>>	24
3.5 <<Dokumentace>>.....	26
3.6 <<Stáhnout celý korpus>>	27
4 Vyhledávací funkce v platformě MERLIN	28
4.1 <<Vytvořit subkorpus>>.....	29
4.2 <<Jednoduché vyhledávání>>	32
4.3 <<Pokročilé vyhledávání>>	34
4.4 <<Statistika>>	41
5 Nápověda	45
5.1 Uživatelský manuál	45
5.2 Seznámení se s platformou MERLIN: video instruktáž.....	45
5.3 Navigace rozhraním MERLIN: manuál s ukázkami	46
5.4 Jak používat platformu MERLIN – příklady užití	46
5.5 Porozumění výstupu <<Pokročilé vyhledávání>>.....	46
5.6 Glosář.....	48
5.7 Časté otázky.....	48
5.8 Kontakt.....	48
Literatura	49

Přehled tabulek/grafiky

1. část: Základní informace	
Tabulka 1	Příklad stupnice a deskriptorů SERR: „Rozsah všeobecných zanlostí jazyka“ (CoE 2001: 110)
Obrázek 1	Strukturální výstavba anotací v platformě MERLIN
Tabulka 2	Příklad cílové hypotézy 1 (TH1)
Tabulka 3	Příklad jevů nezahrnutých do TH1
Tabulka 4	Příklad cílové hypotézy 2 (TH2)
Tabulka 5	Anotace rysů žakovského jazyka (tagy a definice)
2. část: Průvodce pro uživatele	
Obrázek 2	<<dokumentace>>
Obrázek 3	<<MERLIN: v praxi>>
Obrázek 4	<<MERLIN: v praxi>>, <<používání platformy MERLIN pro výuku jazyka>>
Obrázek 5	<<MERLIN: výzkum>>
Obrázek 6	<<korpus MERLIN>>
Obrázek 7	Testové úlohy (<<MERLIN: korpus>>, němčina – ukázka)
Obrázek 8	Detail popisu testové úlohy (<<MERLIN: korpus>>)
Obrázek 9	Počet textů na jednotlivých úrovních SERR a počet hodnocení pro každou úroveň SERR (<<MERLIN: korpus>>)
Obrázek 10	Korpus MERLIN v číslech: počet textů s cílovou hypotézou 1, 2 a anotací chyby 1, 2 (<<MERLIN: korpus>>)
Obrázek 11	<<MERLIN: anotace>>
Obrázek 12	Struktura anotací v platformě MERLIN, celý korpus (<<MERLIN: anotace>>)
Obrázek 13	Seznam anotačních tagů s příklady (<<MERLIN: anotace >>)
Obrázek 14	<<Dokumentace>>
Obrázek 15	<<Stáhnout celý korpus>>
Obrázek 16	Sekce vyhledávání
Obrázek 17	Rozhraní pro vyhledávání <<Vytvořit subkorpus>>
Obrázek 18	Specifikace parametrů subkorpusu, rysy žakovského jazyka (<<Vytvořit subkorpus>>)
Obrázek 19	Příklad výstupu (<<Vytvořit subkorpus>>)
Obrázek 20	Příklad výstupu <<Zobrazit žakovský text a TH>> ve výstupu <<Definovat subkorpus>>
Obrázek 21	Příklad výstupu <<Zobrazit informace o autorovi textu a hodnocení>> ve výstupu <<Vytvořit subkorpus>>
Obrázek 22	Rozhraní pro vyhledávání <<Jednoduché vyhledávání>>
Obrázek 23	Příklad výstupu <<Jednoduché vyhledávání>>
Obrázek 24	Příklad výstupu <<Informace o autorovi textu a hodnocení>> ve výstupu <<Jednoduché vyhledávání>>
Obrázek 25	Rozhraní pro vyhledávání <<Pokročilé vyhledávání>>
Obrázek 26	Definování tagů slovních druhů <<Pokročilé vyhledávání>> (detailně, čeština)
Obrázek 27	Definování <<Kategorie rysu>> v <<Pokročilém vyhledávání>> (detailně, gramatika)
Obrázek 28	Příklad vyhledávání lemma, <<Pokročilé vyhledávání>>
Obrázek 29	Příklad výstupu vyhledávání lemma, <<Pokročilé vyhledávání>>
Obrázek 30	Příklad výstupu vyhledávání lemma, skrytá metainformace vlevo <<Pokročilé vyhledávání>>
Obrázek 31	Upravit rozsah kontextu pro lemma ve výstupu <<Pokročilé vyhledávání>>
Obrázek 32	Zobrazit automatické anotace ve výstupu <<Pokročilé vyhledávání>>
Obrázek 33	Rozhraní pro vyhledávání <<Statistika>>
Obrázek 34	Příklad výpočtu absolutní frekvence výskytu (všechny gramatické chyby, němčina), <<Statistika>>
Obrázek 35	Výstup příkladu výpočtu absolutní frekvence výskytu (všechny gramatické chyby, němčina), <<Statistika>>
Obrázek 36	Všechny složky příkladového výpočtu frekvence výskytu (Obrázek 34-35) ve výstupu <<Pokročilé vyhledávání>>
Obrázek 37	Nápověda
Obrázek 38	Nápověda ve výstupu <<Pokročilé vyhledávání>>
Obrázek 39	Funkce nápovědy <<Vysvětlivky k tomuto výstupu >> ve výstupu <<Pokročilého vyhledávání>>
Obrázek 40	Seznam zkratk používaných ve výstupu <<Pokročilé vyhledávání>>

ČÁST 1: Základní informace

1 Projekt MERLIN – cíl a motivace

Úvod

Společný evropský referenční rámec pro jazyky (SERR) je jedním z nejdůležitějších nástrojů pro výuku jazyků a jazykovou certifikaci v Evropě. Jeho jádro tvoří obecně známý systém úrovní SERR, jehož škála je názorně popsána. Jeho přínos však zdaleka není vnímán pouze v oblasti standardizace a rozvoje výuky jazyků a koncipování jazykových testů. Přestože je dnes používání škály úrovní dle SERR velmi rozšířené – snad neexistuje jazykový test, učební plán či učebnice jazyka, které by této škály nevyužívaly – v mnoha případech se tato škála neopírá o žádné příklady z reálných dat žáků, kteří se učí jazyky. Situace je ještě problematičtější u jazyků jiných než angličtina (srov. např. Fulcher 2004, Hulstijn 2007, North 2000, Wisniewski 2014).

Projekt *MERLIN: Multiligvální platforma pro evropské referenční úrovně: mezijazykový výzkum v kontextu* má za cíl zlepšit současný stav a nabídnout ilustrativní příklady a validaci pro systém úrovní SERR. Projekt MERLIN (2012–2014) byl spolufinancován Evropskou unií v rámci Programu celoživotního vzdělávání, č. 518989-LLP-1-2011-1-DE-KA2-KA2MP. Cílem projektu MERLIN je výzkum a rozvoj empirických základů škály SERR pomocí vytvoření korpusu psaných textů v češtině, němčině a italštině jako druhého jazyka (srov. např. Wisniewski et al. 2013).

Výchozí pozice: škála SERR

Škála SERR má být obecně použitelná pro všechny evropské jazyky. Proto musel být popis jednotlivých úrovní zformulován na obecném základě (viz příklad níže):

B1	<i>Má dostačující vyjadřovací prostředky k tomu, aby popsal(a) nepředvídatelné situace, aby v rozumné míře dostatečně přesně postihl(a) podstatu myšlenky nebo problému a aby se vyjadřoval(a) o abstraktních či kulturních tématech, jako jsou hudba a film.</i>
	<i>Má dostačující vyjadřovací prostředky k tomu, aby se domluvil(a) a aby se vyjadřoval(a) pomocí své slovní zásoby jen s určitou mírou zaváhání a optických jazykových prostředků v rámci tematických okruhů jako jsou rodina, koníčky a zájmy, práce, cestování a aktuální události, ale omezená slovní zásoba způsobuje opakování a občas dokonce formulační obtíže.</i>
A2	<i>Má základní jazykový repertoár, který mu/jí umožňuje, aby se vyjadřoval(a) v rámci každodenních předvídatelných situací, ačkoli bude muset ve většině případů zjednodušit obsah sdělení a hledat slova.</i>
	<i>Ovládá užívání krátkých každodenních výrazů, pomocí nichž vyjadřuje jednoduché potřeby konkrétní povahy: osobní data, denní režim, potřeby a požadavky, požádání o informace. Umí používat základní typy vět a komunikovat o sobě a jiných lidech, o tom, co dělají, o místech, o vlastnictví atd. pomocí pamětně osvojených frází, skupin několika slov a formulací. Má omezený repertoár krátkých pamětně osvojených frází, které se týkají předvídatelných základních jazykových situací; v neobvyklých situacích dochází k častým selháním v jazykové komunikaci a nedorozumění.</i>
A1	<i>Má pouze základní rozsah jednoduchých výrazů týkajících se osobních dat a potřeb konkrétní povahy.</i>

Tabulka 1: Příklad stupnice a deskriptorů SERR: „Rozsah všeobecných zanlostí jazyka“ (CoE 2001: 110)

Pro platformu MERLIN byly použity stupnice z 5. kapitoly SERR (“komunikativní jazykové kompetence”): (rozsah všeobecných znalostí jazyka | gramatická správnost | rozsah slovní zásoby | ovládnání slovní zásoby | pravopis | koherence a koheze | sociolingvistická přiměřenost). SERR je k dispozici ke stažení na webových stránkách Rady Evropy.¹

Ilustrativní popis úrovní SERR

Nicméně se dospělo k závěru, že bude zapotřebí použít ještě jiné jazykově-specifické ilustrace jednotlivých deskriptorů. Ostatně, již od r. 2001 samotná Rada Evropy podporuje myšlenku vytvoření doplňkových nástrojů k SERR, které by lépe vysvětlovaly rysy v jednotlivých jazycích. Jednou z takových iniciativ bylo i vydání publikace *Reference Level Descriptions* (Popisy referenčních úrovní) pro národní a regionální jazyky.

V současnosti převládá silná tendence zakládat tyto popisy na korpusech žákovského jazyka, jako např. korpus angličtiny (www.englishprofile.org), ale také italštiny (Spinelli/Parizzi 2010) a norštiny (Carlsen 2013).

Platforma MERLIN se také snaží ilustrovat úrovně SERR pro dané jazyky, od ostatních korpusů se však liší tím, že jako vůbec první používá vícejazykový přístup. To znamená, že pokrývá tři jazyky z různých jazykových rodin (slovanské, germánské a románské jazyky) a nabízí mezijazykové srovnání.

Platforma MERLIN se zároveň od jiných platforem liší tím, že jde o didaktickou on-line platformu, která svým uživatelům poskytuje přístup k úplným textům, testovým úlohám a širokému spektru lingvistických anotací a anotací chyb zcela zdarma. Platforma MERLIN má také ambici přispět k validitě úrovní SERR.

Validace deskriptorů úrovní SERR

Sestavení stupnic deskriptorů SERR Radou Evropy (CoE 2001; North 2000; Schneider/North 2000) vedlo k významnému rozvoji v oblasti standardizace a transparency výuky a učení se jazyka a koncipování jazykových testů. Referenční úrovně SERR hrají důležitou roli v životech žáků, kteří se učí cizí jazyk. Přesto existuje jeden aspekt, který pořád není dostatečně pochopen a tím je empirická validita škály SERR (Fulcher 2004; Hulstijn 2007). Pokud mají škály popisovat nebo hodnotit úroveň žákovského jazyka, musí odrážet skutečné výstupy těchto žáků (Alderson 1991).

Jelikož se čím dál tím více využívá referenčních úrovní SERR v kontextech, kdy interpretace škály SERR může mít zásadní vliv na život jedince (např. přijímací testy na univerzity, testy pro účely migrace), je obzvláště důležité, aby tato škála odrážela skutečnou úroveň žákovského jazyka. V této oblasti bylo provedeno jen velmi málo výzkumů (srov. např. Alderson et al. 2006; Alderson 2007; Fulcher 2004; Hulstijn 2007; Hulstijn et al. 2010; Little 2007; Wisniewski 2013, 2014).

¹ http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

Kalibrace škály SERR je založena na představách jejích uživatelů o cizojazyčné kompetenci, které se odrážejí v hodnoceních. Přesto není jasné, do jaké míry tato hodnocení souhlasí s popisem úrovní škály (Arras 2010; Eckes 2008; Pollitt/Murray 1996; Vaughan 1991). Během vytváření škály SERR totiž neproběhly žádné analýzy žákovského jazyka, které by umožnily empirickou validaci. Projekt MERLIN chce přispět k takovému výzkumu validity škály SERR. Lingvistické koreláty k obsahu stěžejní škály 5. kapitoly byly operacionalizovány a jsou dohledatelné v rozhraní platformy.

Výzkum v oblasti automatického zpracování jazyka

Korpus platformy MERLIN obsahuje cenná data pro vytváření a evaluaci nástrojů automatického zpracování jazyka pro žákovský jazyk (Meurers 2012). Korpus a metainformace o žácích a hodnoceních v něm obsažené slouží jako okamžitá podpora pro výzkum automatického rozpoznávání mateřského jazyka, což umožňuje posunout výzkum v této oblasti dál než dosavadní zájmu pouze o žáka angličtiny. Podobným způsobem se korpus využívá i pro automatické hodnocení jazykové dovednosti v němčině (Hancke 2013). Korpus MERLIN také poskytuje bohatě anotované žákovské materiály, které mohou sloužit pro vytváření a adaptaci nástrojů automatického zpracování jazyka a vytváření aplikací, díky kterým se žáci mohou zlepšit v oblasti slovní zásoby, koherence jazyka, pravopisu a gramatické přesnosti.

2 Metodologie práce

2.1 Sběr dat

Texty v platformě MERLIN pocházejí ze subtestů psaní standardizovaných testů vysoké kvality vztahujících se k SERR, a to z telc Fankfurt nad Mohanem (italské a německé testy, www.telc.net) a z ÚJOP v Praze (české testy, www.ujop.cuni.cz). Obě zkušební instituce prošly auditem asociace ALTE (www.alte.org). Testové úlohy byly používány do roku 2013 a nyní jsou uživatelům k dispozici na platformě zdarma.

2.2 Transkripce

Ručně psané žákovské texty byly zkušebními institucemi (telc a ÚJOP) transkribovány v xml editoru (xml mind©). Při transkripci postupovali pracovníci podle transkripčního návodu, který je uveden v rozhraní platformy, a reliabilita transkribovaných textů byla následně ověřena nejdříve na vzorku 5 % všech textů pro každou úroveň SERR. Protože bylo detekováno mnoho transkripčních chyb, bylo nutné, aby téměř všechny texty nakonec prošly revizí.

V rámci transkripčního postupu byly vytvořeny tagy (inline annotation) pro základní textové rysy, jako např. nečitelné nebo nejasné pasáže části textu, cizí slova, emotikony, obrázky,

odstavce, zkopírovaná slova ze zadání či oslovení. Součástí transkripčního procesu bylo i např. zanonymizování jmen a míst, při kterém se postupovalo podle zmíněných pravidel. Transkribovaný materiál se stal základem pro vytváření anotací (viz níže). Transkripční postupy jsou k dispozici na rozhraní platformy MERLIN pouze v němčině v sekci <<dokumentace>>.

2.3 Opětovné hodnocení

Původní testy, ze kterých byly vyňaty texty pro platformu MERLIN, byly ohodnoceny určitým počtem bodů, který byl následně vážen podle toho, jaká důležitost byla v rámci testu připsána subtestů psaní. Následně bylo rozhodnuto o celkovém hodnocení testu, buď prospěl, nebo neprospěl. Pro platformu MERLIN musel být postup hodnocení nevyhnutně jiný: cílem bylo dát tyto texty do přímého vztahu ke stupnicím z 5. kapitoly SERR pro komunikativní jazykovou kompetenci. Z toho důvodu byly všechny texty opětovně ohodnoceny nezávislými profesionálními hodnotiteli.

Reliabilita opětovného hodnocení byla ověřena klasickou metodologií pomocí teorie CTT (Classical Test Theory) a analýzou multiparametrickým Raschovým modelem. Druhá ze zmíněných metod je pravděpodobnostní statistická metoda, která se často používá pro koncipování jazykových testů a která umožňuje korekce tendencí hodnotitelů, např. shovívavost/přísnost a díky níž je možné pro každý text vypočítat adekvátní průměr hodnocení. Vnitřní konzistence hodnotitelů (intra-rater reliabilita) i shoda mezi nimi (inter-rater reliabilita) byly v platformě MERLIN obecně na vysoké úrovni, kromě pár výjimek u italských textů. Proto se celý proces opětovného hodnocení musel pro italské texty zopakovat, než se dospělo k uspokojivé kvalitě hodnocení. Více informací je možné najít v Technické zprávě (Technical Report, viz sekce <<dokumentace>>).

Pro platformu MERLIN byla použita holistická škála, založena na škále SERR pro rozsah všeobecných znalostí jazyka a také analytická hodnotící tabulka s následujícími kritérii hodnocení: pravopis | gramatická správnost | rozsah slovní zásoby | ovládání slovní zásoby | koherence a koheze | sociolingvistická přiměřenost, a to v rozpětí úrovní A1 až C2. Oba nástroje jsou k dispozici ke stažení v sekci <<dokumentace>> platformy MERLIN. Adekvátní průměr je vypočítán na základě holistické škály. Když si uživatel vytvoří svůj vlastní korpus, založený na úrovních SERR, budou i jednotlivé úrovně založeny na hodnocení adekvátním průměrem. V tomto rozhraní má uživatel rovněž možnost přístupu k hodnotícím profilům s původním hodnocením zmíněných kritérií hodnocení.

Upozorňujeme, že mnoho z těchto testů bylo původně pro žáky buď příliš náročných, nebo neadekvátně jednoduchých. Proto se pro platformu MERLIN zavedlo rozlišení mezi úrovní SERR testu a úrovní/-ěmi SERR hodnocení (viz také obrázek 9 níže), které se nemusí vždy shodovat. Každý typ úrovně je možné v platformě vyhledat zvlášť (např. je možné vyhledat české texty z testů na úrovni B2, které ale získaly pouze hodnocení úrovně B1).

2.4 Manuální anotace

Anotace je jednou ze základních prvků projektu MERLIN. Platforma MERLIN používá dva typy anotací: **cílové hypotézy** a anotace **rysů žákovského jazyka**. Tam, kde to bylo možné, byly také použity **automatické postupy** (viz 1.4), ale většina anotací byla vytvořena **manuálně**.

Anotace byly seřazeny do 2 bloků:

Anotace, které lze použít pro celý korpus



Cílová hypotéza 1: pravopisně a gramaticky správná verze žákovského textu (TH1)



Rysy žákovského jazyka: pravopis a gramatika (anotace chyby 1, EA1)

Další anotace, které lze použít pro jádrový korpus (A2/B2)



Cílová hypotéza 2 (TH2): akceptovatelný žákovský text



Rysy žákovského jazyka: slovní zásoba, pragmatika, sociolingvistika, srozumitelnost (anotace chyby 2, EA2)

Obrázek 1: Strukturní výstavba anotací v platformě MERLIN

2.4.1 Cílové hypotézy (TH)

Výzkumy ukázaly, že vytváření anotací pro žákovský jazyk je složitý proces, který občas vyžaduje hlubší přístup. Hodnocení každého jevu žákovského jazyka vyžaduje mentální interpretaci anotátora. Aby byla zachována transparentnost, koherence a reliabilita anotací, je vhodné si nejdřív stanovit „cílovou hypotézu“ (TH), tj. vytvořit správnou verzi žákovského jazyka, na níž pak může stavět anotace chyby (Reznicek/Lüdeling et al. 2012). Cílové hypotézy jsou potřebné také pro úspěšnou implementaci mnoha automatických analýz (Díaz-Negrillo et al. 2010, Hirschmann et al. 2009). Třetím aspektem je, že cílové hypotézy mohou pomoci budoucím uživatelům platformy MERLIN porozumět anotacím. Projekt MERLIN spolupracuje s projektem Falko² (Humboldtova univerzita v Berlíně), což je jedna z mála korpusových iniciativ zabývajících se cílovými hypotézami a nabízející volný přístup k datům. Platforma MERLIN používá dva typy cílové hypotézy 1 (TH1 a TH2), které budou v následujícím textu krátce vysvětleny.

² <https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko>

Cílová hypotéza 1

Cílové hypotézy pro ortografické a gramatické chyby (TH1) byly vytvořeny pro celý korpus platformy MERLIN. Od anotátora se u TH1 požaduje, aby text upravil v nejmenší možné míře a vytvořil tak gramaticky a ortograficky správnou verzi původního žakovského textu (*minimální cílová hypotéza*). V následující grafice je uveden příklad:

Žakovský text	Ich	Habe	Seit	5	Jahren	In	meinen	Heimatland	Deutsch	gelernt	(...)
TH1	Ich	Habe	Seit	5	Jahren	In	meinem	Heimatland	Deutsch	gelernt	(...)
TH1Diff							CHA				
Gloss TH1	<i>I</i>	<i>Have</i>	<i>Since</i>	<i>5</i>	<i>years</i>	<i>in</i>	<i>my</i>	<i>home country</i>	<i>German</i>	<i>studied</i>	<i>(...)</i>

Tabulka 2: Příklad cílové hypotézy 1 (TH1)

Následující příklad od stejného žáka znázorňuje, že chyby v jiných lingvistických oblastech pro TH1 nebyly brány v potaz. Je to z obsahových a technických důvodů.

Žakovský text	Ich	Habe		Srilankische	Aufenthalts	und	Reise	Spass	.
TH1	Ich	Habe	Einen	sri-lankischen	Aufenthalts-	und	Reisespass		.
TH1Diff			INS	CHA	CHA		MERGE		
Gloss TH1	<i>I</i>	<i>Have</i>	<i>A</i>	<i>Sri Lankan</i>	<i>residence</i>	<i>and</i>	<i>travel fun</i>		.

Tabulka 3: Příklad jevů nezahrnutých do TH1

Zatímco se ortografické chyby (např. psaní velkých písmen, hranice slova, chybějící spojovník) a gramatické chyby (např. člen) v TH1 opravují, lexikálně chybná forma *Reisespass („zábava cestování“) místo Reisepass („cestovní pas“) nebyla nahrazena jiným lexémem.

Cílová hypotéza 2

Jako druhý stupeň se cílové hypotézy 2 (TH2) vztahují k sociolingvistickým, lexikálním a pragmatickým odchylkám od forem, které jsou běžně očekávány od rodilého mluvčího. TH2 má tedy za cíl vytvořit *akceptovatelnou* verzi původního žakovského textu. U TH2 je brán v úvahu kontext. TH2 je rozšířená TH1.

Aby se dospělo k spolehlivému rozhodnutí, je pro TH2 zapotřebí subjektivnější přístup než u TH1, a proto je tento proces také problematičtější než u TH1. To je i jedním z důvodů, proč jsou obě vrstvy od sebe odděleny. Následující tabulka názorněji vysvětluje rozdíl mezi TH1 a TH2:

Žakovský text	Ich	Habe		Srilankische	Aufenthalts	und	Reise	Spass	.
TH1	Ich	Habe	Einen	sri-lankischen	Aufenthalts-	und	Reisespass		.
TH1Diff			INS	CHA	CHA		MERGE		
TH2	Ich	Habe	Einen	Sri-lankischen	Aufenthalts-	und	Reisepass		.
TH2Diff			INS	CHA	CHA		MERGE/CHA		
Gloss TH2	<i>I</i>	<i>Have</i>	<i>A</i>	<i>Sri Lankan</i>	<i>residence</i>	<i>and</i>	<i>travel passport</i>		.

Tabulka 4: Příklad cílové hypotézy 2 (TH2)

Z tabulky je zřejmé, že na úrovni TH2 pro stejnou větu, pro kterou byla vytvořena TH1, byla potřebná nová lexikální anotace („Reisepass“ („cestovní pas“) místo *Reisepass („zábava cestování“)).

2.4.2 Anotace rysů žakovského jazyka

Jedním z důležitých principů budování platformy MERLIN je také nahlížení na žakovský jazyk jako na autonomní systém, který nelze dostatečně popsat pouze tagy chyby, jež se soustřeďují na chybějící elementy. Je zde potřeba zdůraznit, že platforma MERLIN sice obsahuje mnoho tagů chyby, což se týká hlavně úrovně EA2, avšak zároveň obsahuje tagy, které zachycují nechybové jevy, jako např. realizace promluvového aktu žádosti.

Anotační schéma platformy MERLIN tedy představuje výběr smysluplných, validních a realizovatelných rysů („tagů“), které jsou manuálně anotované a podpořené prací týmu počítačových lingvistů projektu MERLIN. Platforma obsahuje tyto rysy:

G_Gramatika

G_Agr	shoda podmětu a přísudku
G_Art	člen
G_Clit	IT: příklonka
G_Conj	spojka
G_Inflect_inexist	neexistující flexe (podstatná jména, přídavná jména, slovesa)
G_Morphol_wrong	chybná flexe (podstatná jména, zájmena, přídavná jména)
G_Neg_negdoub	ČJ: dvojí negace
G_Neg_neggen	negace obecně
G_POS	chybný slovní druh
G_Prep	předložka
G_Refl_pronrefl	zvrtné zájmeno
G_Refl_pronreflposs	ČJ: přivlastňovací zvrtné zájmeno
G_Valency_complnumb	valence slovesa: počet nezbytných doplňků
G_Verb_asp	sloveso: slovesný vid (ČJ+IT)
G_Verb_compl	konjugace slovesa (morfologie)
G_Verb_main	plnovýznamové sloveso
G_Verb_md	sloveso: způsob
G_Verb_tns	sloveso: čas
G_Verb_vc	sloveso: slovesný rod
G_Wo_womaincl	slovesled v hlavní větě
G_Wo_wosubcl	slovesled ve vedlejší větě

O_Pravopis

O_Abbrev	zkratka
O_Apostr	NJ+IT: apostrof
O_Capit	psaní velkých písmen
O_Graph_act	ČJ+IT: diakritická znaménka
O_Graph_graphgen	chybný přepis

O_Graph_trans	záměna písmen
O_Punct	interpunkce
O_Wordbd	hranice slov v písmu

G_Srozumitelnost

G_Intelltxt	srozumitelnost textu
G_Intells	srozumitelnost věty

V_Slovní zásoba

V_FS	ustálené slovní spojení
V_Sequence	nesrozumitelná část, kdy příčinou nesrozumitelnosti je
_lexgrammerr_incompr	lexikální/gramatická chyba/chyby
V_form_nonexist	neexistující forma (slovo nebo slovní spojení)
V_FS_form_incompr	ustálené slovní spojení: omezená srozumitelnost
V_sendenot	sémantická chyba: denotace (slovo nebo slovní spojení)
V_semcon_att	sémantická chyba: konotace (subjektivní vnímání významu slova), (slovo nebo slovní spojení)
V_Word_semimprec	sémantická chyba: přesnost významu (slovo nebo slovní spojení)
V_Wordform_deriv	chybné tvoření slov: derivace
V_Wordform_comp	chybné tvoření slov: kompozice
V_FS_form	ustálené spojení slov: chybný tvar slova

C_Koherence/Koheze

C_Con_accur	přesnost užití spojky
C_Coh_jump	skoky v obsahu
C_Coh_ref	odkaz
C_Coh_txtstruct	prostředek metakomunikace

S_Sociolingvistická přiměřenost

S_Txt_grfw	pozdrav/rozloučení
S_Txt_opcl	úvodní/závěrečná formulace
S_Form_gen	nevhodné užití prostředků (formálnost)
S_Form_addr	nevhodné oslovení (formálnost)
S_Var_clit	IT: lexikalizovaná příklonka (verbi procomplementari)
S_Var_duppron	IT: nadbytečnost osobních zájmen
S_Var_synstr	IT: srozumitelné syntaktické struktury
S_Var_che	IT: 'che polivalente'
S_Var_woweil	NJ: chybný pořádek slov ve větě po spojce 'weil'
S_Var_partik	NJ: modální částice

P_Pragmatika

P_Pol_dir	zdvořilost - nezdvorné použití imperativu
P_Request_direct	přímá žádost
P_Request_indirect	nepřímá žádost

Tabulka 5: Anotace rysů zákovského jazyka (tagy a definice)

Uvedené tagy byly pro projekt vybrány ze seznamu jazykových rysů a indikátorů, které byly v rámci projektu shromážděny. Možné anotace byly **shromážděny z následujících oblastí**:

- 1) pohled uživatelů platformy (na základě zkoumání uživatele a rozborů učebnic a jazykových testů)
- 2) škála SERR (operacionalizací prvků 5. kapitoly, CoE 2001)
- 3) výzkumy (založené na rozsáhlé analýze zdrojové literatury)
- 4) žákovské texty (na základě induktivní analýzy 10 % z celkového počtu žákovských textů)

Uživatelé projektu MERLIN mají k dispozici celou řadu informačních zdrojů, které se týkají anotací:

- použitá **bibliografie** pro výběr smysluplných anotačních tagů pro platformu MERLIN
- **anotační schéma** obsahující anotační tagy, které byly implementovány do platformy MERLIN po praktickém ověření seznamu maximálního počtu anotací (k dispozici **ke stažení**)
- dokumenty jako např. **manuál pro anotátory platformy MERLIN** (anglicky) nebo **dokumentace dalších anotačních řešení**, které poskytují pečlivě vypracované řešení jednotlivých anotačních jevů pro všechny tři jazyky tohoto projektu, či jiné materiály, umístěné na tomto rozhraní (k dispozici **ke stažení**)
- funkce nápovědy rozhraní, např. **seznam anotačních tagů s příklady, ukázky jednotlivých stránek** vysvětlující anotační **vrstvy**, dále **glosář** aj.
- **black book** obsahující zkušenosti s vytvářením anotací, která může pomoci lépe se zorientovat v této oblasti při jakýchkoliv budoucích projektech (k dispozici **ke stažení**)

Anotační tagy doporučené uživateli

Platforma MERLIN má sloužit uživatelům, kteří pracují se stupnicemi SERR. Proto je důležité vytvářet anotace tím způsobem, aby byly co nejužitečnější. Aby mohlo být toto splněno, proběhl **výzkum uživatelů**, jehož podrobnosti jsou uvedeny ve dvou zprávách nacházejících se na rozhraní (<<*dokumentace*>>). První se soustředí na užitečnost obsahových prvků, druhá bere v potaz technickou stránku používání platformy MERLIN. Výzkum přinesl důležité poznání o tom, které anotační tagy by měly být pro platformu použity. Anotační schéma obsahuje informace o jednotlivých typech tagů, které vyplynuly z výzkumu.

Dalším způsobem, jak pochopit potřeby uživatelů platformy, je integrovat do anotačního schématu platformy MERLIN prvky osvojování druhého jazyka L2, které jsou často obsažené v **učebnicích a jazykových** testech. Mezi publikacemi, které prošly analýzou pro platformu MERLIN, byly např. „Tangram“ pro němčinu (Dallapiazza 1998), „Rete!“ pro italštinu (Mezzadri 2000) a „Brána jazyka českého otevřená“ pro češtinu (Hasil 2007). Také analýzy

provedené na jazykových testech institucemi ÚJOP UK a telc přinesly určitá notoricky známá témata, která byla využita pro anotační schéma.

Mezi anotace, které byly odvozeny z těchto analýz, patří např. pravopisné chyby (chyby v psaní velkých písmen, nesprávné užití apostrofu v němčině a italštině) nebo gramatické chyby, jako např. slovesný vid v italštině. V oblasti lexika se opakovaně objevovaly chyby v používání slov se zdánlivě stejným významem, tzv. false friends nebo frazeologismů a tyto poznatky se odrazily v anotačním schématu platformy MERLIN.

Žákovské texty jako zdroj odvozování anotačních tagů

10% všech žákovských textů v korpusu bylo analyzováno ručně. Díky tomuto kvalitativnímu a induktivnímu přístupu bylo možné odhalit množství jazykových jevů, které mělo význam integrovat do anotačního schématu. Mezi takové jevy patří chyby v tvarosloví, různé typy problémů s používáním ustálených slovních spojení a problémy týkající se výběru vhodného rejstříku v rámci formálnosti a zdvořilosti. Zajímavý je fakt, že většina těchto jevů patří do jiných kategorií než gramatika či pravopis.

Anotace založená na výzkumu

Mnoho úsilí bylo také věnováno studiu odborné literatury z různých jazykových oblastí, které platforma MERLIN pokrývá. Pro kompletní rozpravu o jednotlivých tazích zde není prostor, proto uvádíme alespoň krátký přehled. Citované reference jsou mj. k nalezení v bibliografickém seznamu.

Pod anotace **gramatických** jevů patří např. shoda podmětu a přísudku, slovosled, negace, chyby ve slovním druhu aj. Zde je tedy možné vypočítat několik parametrů přesnosti a komplexity textu (viz Lu 2010, 2011; Wolfe-Quintero et al. 1998, Ortega 2003, 2012, Housen/Kuiken 2009). **Pravopis** nepatří mezi oblasti, kterou by se výzkumy detailněji zabývaly. Díky anotacím platformy MERLIN máme přístup k relativní/absolutní frekvenci výskytu různých typů pravopisných chyb, např. interpunkce, psaní velkých písmen či diakritika (Al-Jarf 2009, Bredel 2010, Cook 2005, Granger/Bestgen 2011, Perfetti/Rieben/Fayol 1997, Rimrott/Heift 2008, Sassoon 1995).

Pilotní jádrový korpus projektu MERLIN (EA2)

Pro menší **subkorpus** byly anotovány také jevy z jiných lingvistických oblastí (EA2, anotace chyby 2). Texty v tomto subkorpusu obsahují anotaci TH1 a EA1 a navíc také anotaci TH2 a EA2. V budoucnu by bylo žádoucí tyto pilotní výzkumné anotace překontrolovat, aby bylo možné doplnit anotace TH2 a EA2 pro celý korpus MERLIN. V současné době je tedy třeba, aby uživatelé přistupovali k anotacím v jádrovém korpusu s opatrností.

Anotací tagy pro **slovní zásobu** berou v potaz řadu dimenzí lexikálních dovedností, jako např. správnost užití, hloubka, šířka a sofistikovanost (Nation 2001, 2007, Read 2000). Zde se

platforma MERLIN silně zaměřuje na ustálená slovní spojení (Wray 2002), která hrají specifickou roli při osvojování cizího jazyka (např. Pawley/Syder 1987, Schmitt et al. 2004). Mnoho lexikálních tagů, týkajících se ustálených slovních spojení, nesouvisí s chybovostí, nýbrž se snaží o zachycení struktur specifického pole zájmu během procesu osvojování.

Sociolingvistická kompetence je definována jako „schopnost rozpoznat a produkovat vhodný styl projevu v kontextu“ (Lyster 1994: 263). Pokud jde o anotace EA2 obecně, anotování jazykových jevů, které se týkají této kompetence (subjektivita, reliabilita), je metodologicky náročné. Použité tagy jsou v souladu se strukturou projektu a ne vždy odpovídají tomu, co je obvykle v této oblasti jazyka analyzováno (např. úroveň ovládnutí jazyka a množství jazykového kontaktu nebo diasystematické variování v žákovských textech ve srovnání s variováním v mateřském jazyce, Baker 2010, Baylea 2007, Bayley/Regan 2004, Biber/Finegan 1994, Hudson et al. 2005, Hymes 1974, Mougeaon/Dewaele 2004, Regan et al. 2009, Van Compernelle/Williams 2012, Yu 2012, Zuskin 1992). Platforma MERLIN obsahuje anotace přiměřenosti jazykových forem v rámci formálnosti a zahrnuje tagy, které jsou vytvořené dle typu textu, jako např. pozdravy nebo úvodní a závěrečné formulace. Jednotlivé vybrané jazykově-variační prvky odrážejí výběr struktury, jež se buď nevztahují k psanému jazyku a/nebo stojí na prahu přijatelnosti coby standardní varianty, ale běžně by nebyly akceptovatelné pro typy testových úloh, v nichž se objevují v textech platformy MERLIN.

V oblasti **pragmatiky** je anotována **koherence/koheze**, a to hlavně označením užití metakomunikačních prostředků, také nezávisle na správnosti, konektorů či anotací problémů v referenci (Bachmann 2002, Halliday/Hasan 1976, 1989, Castro 2004, Carlsen 2010, Chiang 2003, Cornish 2009, Fabricius-Hansen 2005, Louwense/Graesser 2004, McNamara et al. 1996, McNamara/Kintsch 1996, Spooren/Sanders 2008). Mluvní akt žádosti je také anotován (Al-Gahtani/Roever 2012, Bardovi-Harlig 2013, Barron 2003, Blum-Kulka 1987, 1991, Blum-Kulka/Olshtain 1984, Cho 2005, Held 1995, Nuzzo 2007, Trosborg 1995, Veddersen 2007).

Anotace podle SERR

Abychom zjistili, jestli škála SERR odráží žákovský jazyk, je důležité operacionalizovat jejich deskriptory, aniž by se muselo použít hodnocení člověka, kde se často ukázalo, že nebylo založeno na použití hodnotících nástrojů, přestože jsou tyto nástroje považovány za spolehlivé (Eckes 2008, Wisniewski 2010). Jakmile nabydou deskriptory jednotlivých stupnic měřitelnou formu, začne být vztah mezi danou škálou SERR a žákovským jazykem jasnější.

Během tohoto operacionalizačního procesu bylo nutné vyloučit příliš vágní, seberefrenční či subjektivní formulace v popisech úrovní (např. „Dokáže komunikovat s rodilými mluvčími, aniž by je podvědomě a nechtěně bavil(a), iritoval(a), nebo aniž by je přinutil(a), aby se k němu/ní chovali jinak, než by se chovali k rodilému mluvčímu.“ (adekvátnost z hlediska sociolingvistiky, úroveň B2, CoE 2001: 122), ale ani aspekty, které se jasně týkají výhradně mluveného jazyka, se nebraly v potaz (např. „S určitým znatelným úsilím se dokáže účastnit diskuse ve skupině [...]. (adekvátnost z hlediska sociolingvistiky, úroveň B2, CoE 2001: 122))

(srov. Wisniewski 2013, 2014). Pokud však popis úrovně zmiňuje pozdravy, skoky v obsahu, idiomatické výrazy nebo fráze jako vlastnosti konkrétních úrovní SERR, byla využitelnost těchto rysů pro anotační schéma platformy MERLIN ověřena, i když tyto tzv. škálové proměnné nemusí hrát žádnou roli ve výzkumu a je často obtížné je definovat. Tyto anotace umožňují ověřit empirickou relevanci dané škály SERR. Kdyby byl obsah popisu škály dostatečně jasný a spolehlivě pozorovatelný na výkonech žáků, byl by to znak empirické validity. Platforma MERLIN nemůže poskytnout celkovou validaci škály SERR, proto se soustřeďuje na vybrané, smysluplné aspekty.

Přístup k anotacím pomocí platformy MERLIN

Přímý přístup k anotacím je možný díky vyhledávacím funkcím platformy MERLIN (<<pokročilé vyhledávání>>, <<vytvořit subkorpus>>), přičemž každý výskyt vyhledávání je zobrazen v kontextu.

Další možností, jak získat přístup k anotacím, je použít je pro **statistický výpočet**. Nejjednodušším způsobem je spočítat celkový počet tagů, který se objevuje v (sub)korpusu. V mnoha případech je však smysluplnější použít anotace pro výpočet normalizovaných dat (tj. pro platformu MERLIN výpočet na jednu větu, T-složku nebo token). Jako příklad můžeme uvést možnost porovnat průměr výskytu morfologických chyb v českém žákovském textu na úrovni B1 oproti textu na úrovni B2. Normalizovaná data manuálně anotovaných jevů jsou počítána na základě postupu automatické segmentace (viz níže). Údaje frekvence výskytu společně s několika dalšími komplexnějšími údaji o komplexitě textu jsou k dispozici pomocí funkce <<Statistika>>.

2.5 Automatické anotace

Automatické anotace pro korpus MERLIN slouží jako podpora vytváření manuálních anotací a jako způsob, jímž se získává přístup k širšímu spektru lingvistických rysů pro výpočet indikátorů a pro přímé vyhledávání v korpusu.

Automatické anotace v korpusu MERLIN lze rozdělit do čtyř kategorií:

1. Lingvistické jednotky potřebné pro manuální anotaci:
 - a) tokeny
 - b) věty
2. Lingvistické jednotky potřebné pro výpočet statistických údajů:
 - a) t-složky
 - b) spektrum typů vět
3. Lingvistická anotace využívající existující definice a nástroje
 - a) slovní druh
 - b) lemma
 - c) složková a závislostní analýza
4. Lingvistická anotace s definicemi a nástroji, specifickými pro platformu MERLIN
 - a) opakování uvnitř textů
 - b) citace materiálu s testovou úlohou

Použili jsme existující nástroje automatické anotace pro cílové jazyky s cílem rozšířit spektrum dostupných lingvistických anotací o mnohem větší počet, než by bylo možné za použití časově a finančně náročné manuální anotace. Je však důležité si uvědomit, že automatické vytváření anotací pro žákovský jazyk je poměrně náročný proces, protože žákovský jazyk se od cílového jazyka, jak po stránce lingvistické, tak po stránce pravopisné či sémantické, často značně liší.

Pro všechny tři jazyky korpusu MERLIN byly použity následující anotační nástroje:

Pro vytvoření tokenů byl použit tokenizátor pro indoevropské jazyky softwaru LingPipe, jehož výstupy byly následně korigovány ručně. Anotace vět byly vytvořeny pomocí segmentátoru vět OpenNLP. Opakující se jevy byly zachyceny pomocí *Saphre library* na základě automatického vytváření anotací slovních druhů a lemmat.

Pro více informací o specifických jazykových nástrojích, které se používají pro automatickou anotaci, navštivte sekci <<MERLIN: výzkum>>.

2.6 Kontrola kvality

Pro zajištění a kontrolu reliability anotace bylo zavedeno několik opatření. Prvním je, že všechny nástroje (pravidla TH1 a TH2, anotační schéma EA1 a EA2) byly nejdříve otestovány v zkušebním režimu, poté byly upraveny a následně implementovány. Zkušební režim byl rozdělen do dvou fází. Nejdříve byl testován samotný anotační proces, aby bylo možné získat základní představu o konkretizaci a změnách v postupu vytváření anotací a nástrojů. Poté, již během zkušebního režimu, všichni anotátoři jednoho jazykového týmu anotovali určitý počet textů, aby se znova vyzkoušelo, zdali je anotační postup praktický a dostačující a aby bylo možné detekovat případné technické problémy.

Dalším opatřením je, že vytváření veškerých anotací je založeno na jasně definovaném postupu (anotační manuál, viz platforma projektu). Tento postup je obohacen o detailně vypracovaná rozhodnutí ohledně jednotlivých prvků anotace (dokument o dalších anotačních řešeních, viz rozhraní). Třetím opatřením je to, že reliability anotací je ověřována. Reliability anotací byla ověřena na vzorku 5 % textů pro každou úroveň testu pro cílové hypotézy 1 a 2 a anotaci chyby 1 a 2. Pro ověřování byly použity různé metody:

Pro kvalitativní přístup byla polovina těchto souborů anotována nezávisle na sobě všemi anotátory, poté proběhla společná diskuse, jejímž cílem bylo dospět ke konsenzu. Použité texty sloužily během celého anotačního procesu jako referenční texty. Kvalitativní přístup se ukázal jako velice důležitý pro obecné porozumění anotačnímu schématu. Druhá polovina souborů, u kterých se ověřovala reliability, byla anotována všemi anotátory, aniž by věděli, o které texty se jedná (double-blind procedure). U těchto souborů pak byla ověřena reliability jak kvalitativně, tak kvantitativně.

Část 2: Průvodce pro uživatele

3 Sekce dokumentace platformy MERLIN

Makrostruktura volně přístupné platformy MERLIN je uspořádána do **části dokumentace** (umístěna vertikálně na levé straně obrazovky rozhraní, viz Obrázek 2) a **části vyhledávání** (horizontální vyhledávací pole, viz Obrázek 23). V kapitole č. 3 je vysvětlen nejdůležitější obsah a funkce sekce dokumentace.

The screenshot shows the MERLIN platform interface. On the left is a vertical navigation menu with items: 'MERLIN: v praxi', 'MERLIN: výzkum', 'MERLIN: korpus', 'MERLIN: anotace', 'Dokumentace', and 'Ke stažení'. The main content area is titled 'Úvod | Vyhledávání | O projektu MERLIN'. It features a search bar with 'Vyhledávání slova' and a dropdown menu for 'Vyhledávat v:' set to 'žakovský text'. Below the search bar are two informational boxes: 'Korpus MERLIN' and 'Využijte korpus MERLIN ...'. The 'Korpus MERLIN' box states that the platform offers access to 2,286 texts written by students of Czech, Italian, and German, and that they are part of standardized tests within the SERR framework. The 'Využijte korpus MERLIN ...' box mentions that it helps in gaining better knowledge of the SERR framework for languages.

Obrázek 2: <<dokumentace>>

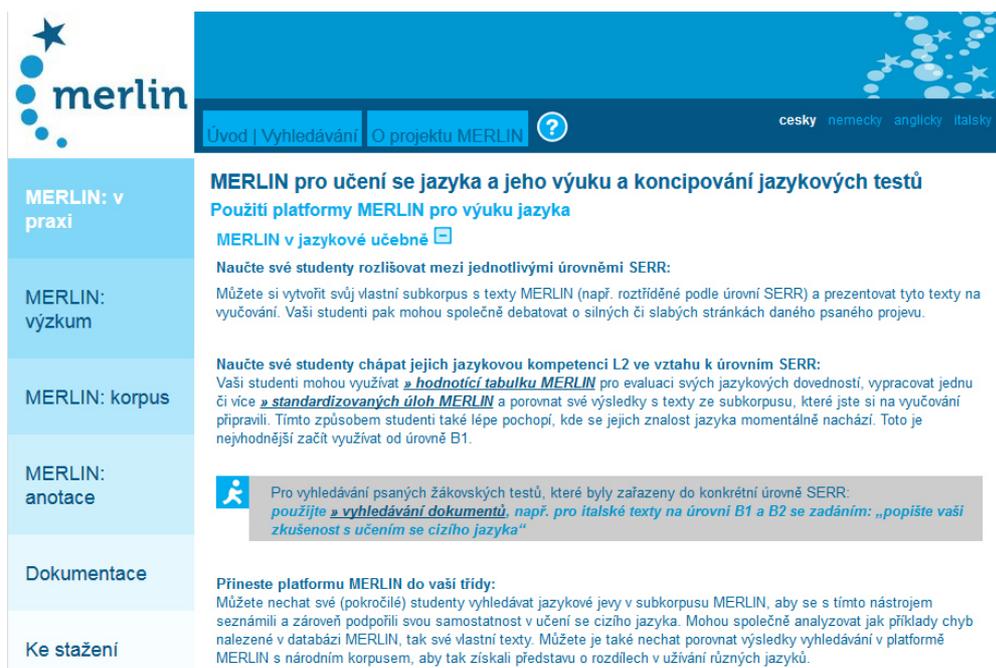
3.1 <<MERLIN: v praxi>>

The screenshot shows the MERLIN platform interface with the 'MERLIN: v praxi' section selected in the navigation menu. The main content area is titled 'Úvod | Vyhledávání | O projektu MERLIN'. The main heading is 'MERLIN pro učení se jazyka a jeho výuku a koncipování jazykových testů'. Below this, there are several sub-sections: 'Použití platformy MERLIN pro výuku jazyka', 'MERLIN v jazykové učebně', 'Naučte své studenty rozlišovat mezi jednotlivými úrovněmi SERR:', 'Praktická pomoc pro autory učebních materiálů', 'Seznamte se se zásadními aspekty v procesu učení se jazyka, jako např. jak studenti používají kolokace, slovesný vid a způsob apod., a získejte vhodné příklady, které použijete ve svých učebních materiálech.', 'Vytváření sylabů a kurikul', 'Většina sylabů, kurikul a dokonce státem požadovaných kritérií pro úroveň vzdělávání v Evropě se řídí referenčním rámcem SERR.', and 'Použití platformy MERLIN pro koncipování jazykových testů'.

Obrázek 3: <<MERLIN: v praxi>>

V sekci <<MERLIN: v praxi>> naleznete obecné informace o možnostech využití platformy MERLIN v různých profesních sférách. V této sekci se seznámíte se základním konceptem tohoto projektu.

Rozbalením jednotlivých kapitol (kliknutím na ikonu ) se můžete dozvědět více např. o tom, jak lze využívat platformu MERLIN pro výuku jazyků nebo pro koncipování jazykových didaktických materiálů.

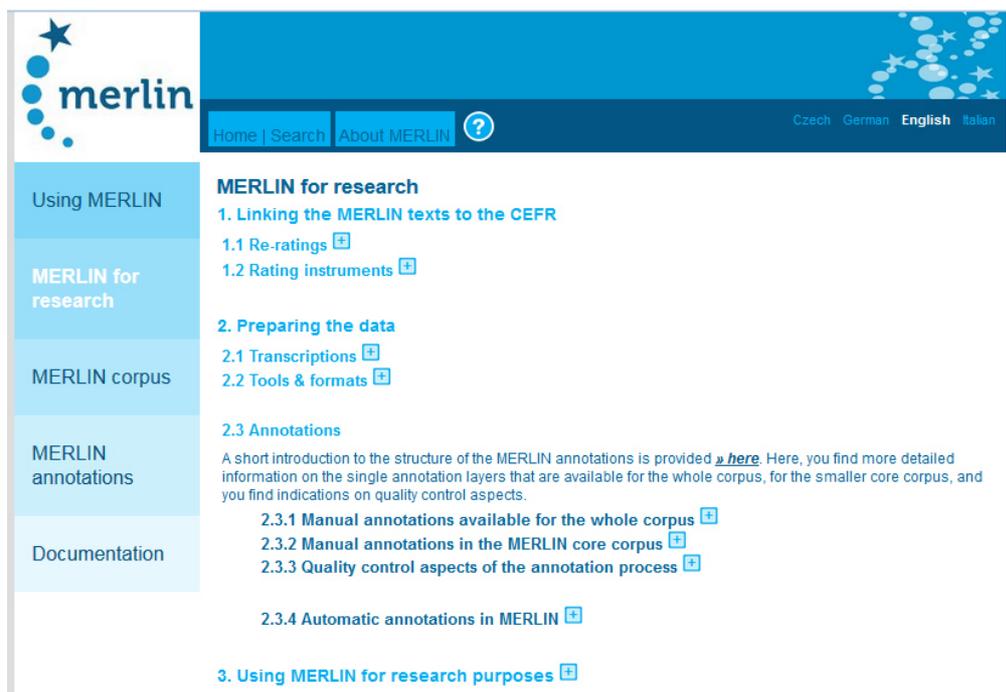


The screenshot shows the MERLIN website interface. On the left is a sidebar with navigation options: 'MERLIN: v praxi', 'MERLIN: výzkum', 'MERLIN: korpus', 'MERLIN: anotace', 'Dokumentace', and 'Ke stažení'. The main content area is titled 'MERLIN pro učení se jazyka a jeho výuku a koncipování jazykových testů'. It includes a sub-header 'Použití platformy MERLIN pro výuku jazyka' and a link 'MERLIN v jazykové učebně'. The text describes how to use the platform for teaching, mentioning SERR levels and the 'hodnotící tabulku MERLIN'. A search tip box is also visible, suggesting search terms like 'vyhledávání dokumentů' for specific SERR levels.

Obrázek 4: <<MERLIN: v praxi>>, <<používání platformy MERLIN pro výuku jazyka>>

3.2 <<MERLIN: výzkum>>

Sekce <<MERLIN: výzkum>> je určena pro zájemce o základní informace o projektu a o jeho dalších aspektech. Pokaždé, když si budete prohlížet rozhraní MERLIN a budete se chtít dozvědět více o důvodech pro vznik tohoto projektu, můžete se podívat do sekce <<MERLIN: výzkum>>.



Obrázek 5: <<MERLIN: výzkum>>

Z obrázku č. 5 je zřejmé, že platforma obsahuje množství informací o **vztahu textů platformy MERLIN k SERR**. Všechny podstatné dokumenty, jako např. hodnotící tabulky, testové úlohy, technická zpráva o kvalitě hodnocení aj., lze najít v této sekci.

V této sekci také naleznete informace o **postupu práce s daty** pro platformu MERLIN: můžete se zde dozvědět o provedené transkripci a anotaci a o tom, jakými nástroji a dle jakého schématu byly prováděny. Zároveň zde naleznete informace o kontrole kvality prvků manuální a automatické anotace.

Jsou zde také uvedeny informace o možnostech **využití platformy MERLIN pro zkoumání** validity škály SERR, osvojování druhého jazyka a automatického zpracování jazyka pro žákovský jazyk. K nalezení je zde také **seznam relevantních referencí**.

3.3 <<MERLIN: korpus>>

The screenshot shows the MERLIN website interface. At the top, there is a navigation bar with the MERLIN logo on the left and language options (cesky, nemecky, anglicky, italsky) on the right. Below the navigation bar, there is a main content area with a sidebar on the left and a main text area on the right. The sidebar contains links for 'MERLIN: v praxi', 'MERLIN: výzkum', 'MERLIN: korpus', 'MERLIN: anotace', 'Dokumentace', and 'Ke stažení'. The main text area is titled 'Korpus MERLIN' and contains several sections: 'Korpus MERLIN' (describing the corpus), 'Texty a pracoviště jazykových zkoušek', 'Vztah k SERR - hodnotící tabulka platformy MERLIN', and 'Zadání testů'. The 'Zadání testů' section is highlighted in blue.

Obrázek 6: <<korpus MERLIN>>

Sekce <<MERLIN: korpus>> čerpá hlavně z dat platformy MERLIN. Zde se můžete dozvědět více o použitých testech a zkušebních institucích, stáhnout si hodnotící tabulky, které byly použity a také seznam testových úloh:

Zadání testů

Poskytujeme komplexní přehled testových zadání podle cílového jazyka a úrovně SERR.

Původní úroveň testu se může lišit od té, kterou byl daný text označen po novém hodnocení úrovně.

Zadání testů se řídí » [tabulkou](#), která byla pro tyto účely vytvořena Evropskou asociací pro jazykové testování ALTE » www.alte.org). Tato tabulka obsahuje podrobné informace o zadáních a specifických rysech textů, které se mají použít ke zkoušce, např. o tématu, rejstříku nebo oblasti textu (autor: Olaf Bärenfänger).

» [Obecné informace o popisu zadání](#)

Němčina

- A1
 - [Neformální e-mail: požádejte přítele o pomoc s hledáním bytu](#)
 - [Neformální e-mail: domluvte se s přítelem, že půjdete spolu plavat](#)
 - [Neformální dopis: napište gratulaci k narození dítěte](#)
- A2
 - [Formální dopis: bytovému úřadu](#)
 - [Neformální dopis: poproste přítele, aby vám pohlídal vašeho domácího mazlíčka](#)
 - [Neformální dopis: nabídněte kamarádovi volný lístek](#)
- B1
 - [Neformální dopis: příteli k Novému roku](#)
 - [Neformální dopis: příteli, kterého přijedete navštívit](#)
 - [Neformální dopis: s přáním k narozeninám](#)

Obrázek 7: Testové úlohy (<<MERLIN: korpus>>, němčina – ukázka)

Když kliknete na testovou úlohu, v novém okně se otevře soubor ve formátu pdf, který obsahuje jednak danou úlohu, jednak podrobný popis úlohy ve formě tabulky, kterou zavedla Evropská asociace jazykových testerů ALTE (www.alte.org). Popis testové úlohy poskytuje více informací o délce úlohy, o typu jazyka v očekávané odpovědi, o obtížnosti úlohy a mnoho dalších informací.

i) Task input/prompt		
17	Language of input/prompt	German
18	CEFR level of input/prompt	A1
19	Time permitted or suggested for this task	n.a.
20	Control/guidance	Semi-controlled
21	Content	Specified
22	Genre	Letter
23	Rhetorical function(s) of input	Describing, instructing
24	Imagined audience	Friend
25	Mode of input/prompt	Written
26	Topic or theme of input	House and home/environment
27	Integration of skills for input	Reading

Obrázek 8: Detail popisu testové úlohy (<<MERLIN: korpus>>)

Sekce <<MERLIN: korpus>> také poskytuje informace o dostupných metadatech, jako např. věk, pohlaví nebo mateřský jazyk, najdete v zde i tabulku celkového počtu textů jako pro jednotlivé úrovně testu, tak pro celkové hodnocení podle SERR:

Korpus MERLIN v číslech

Počet textů pro jednotlivé úrovně SERR testu („úroveň testu“) ve srovnání s počtem textů pro jednotlivé úrovně SERR, u kterých byla provedena změna hodnocení („adekvátní průměr“)

	Úroveň testu		Adekvátní průměr	
			A1	1
čeština	A2	111	A2	189
	B1	143	B1	165
	B2	188	B2	81
			C1	2
italština	A1	207	A1	29
	A2	202	A2	378
	B1	201	B1	394
	B2	201	B2	2
němčina	A1	206	A1	57
	A2	209	A2	297
	B1	210	B1	331
	B2	204	B2	293
	C1	204	C1	42
			C2	4
Celkem	2286	2265		

Obrázek 9: Počet textů na jednotlivých úrovních SERR a počet hodnocení pro každou úroveň SERR (<<MERLIN: korpus>>)

A také tabulku s počtem anotačních vrstev, které mají texty platformy MERLIN:

	čeština	němčina	italština
celkový počet textů	442	1033	813
TH1	440	1033	813
EA1	361	752	754
TH2	231	275	154
EA2	198	258	85

Obrázek 10: Korpus MERLIN v číslech: počet textů s cílovou hypotézou 1, 2 a anotací chyby 1, 2 (<<MERLIN: korpus>>)

3.4 <<MERLIN: anotace>>

The screenshot shows the 'Anotace v korpusu MERLIN' page. The header includes the MERLIN logo and navigation links: 'Úvod | Vyhledávání | O projektu MERLIN'. The sidebar on the left lists: 'MERLIN: v praxi', 'MERLIN: výzkum', 'MERLIN: korpus', 'MERLIN: anotace', and 'Dokumentace'. The main content area is titled 'Anotace v korpusu MERLIN' and contains the following sections:

- Obecné informace**
 - [Struktura anotací](#)
 - Data v platformě MERLIN jsou doplněna o anotaci na několika úrovních.
 - [Anotační schéma – rysy žakovského jazyka](#)
 - Pro anotaci rysů žakovského jazyka vytvořil tým projektu MERLIN [anotační schéma](#) ...
- [Seznam rysů žakovského jazyka s příklady](#)

Obrázek 11: <<MERLIN: anotace>>

Tato sekce obsahuje informace týkající se anotací v platformě MERLIN, přičemž je pozornost kladena na vytváření manuálních anotací.

Zde můžete vidět přehled struktury anotací korpusu MERLIN:

Struktura anotací

Data v platformě MERLIN jsou doplněna o anotaci na několika úrovních.

Ačkoliv většina rysů žákovského jazyka musela být anotována manuálně, metoda automatického zpracování přirozeného jazyka byla použita pro **automatické anotace rysů žákovského jazyka** jako tokeny, lemmata, tagy částí promluvy, segmentace na věty nebo na T-složky (T-units).

Anotace v celém korpusu MERLIN

Anotace, které lze použít pro celý korpus



Cílová hypotéza 1: pravopisně a gramaticky správná verze žákovského textu (TH1)



Rysy žákovského jazyka: pravopis a gramatika (anotace chyby 1, EA1)

Hlavní anotace, které lze použít pro celý korpus MERLIN, jsou cílové hypotézy (**cílové hypotézy 1**) a anotace gramatických a pravopisných rysů žákovského jazyka (**anotace chyby 1**).

Obrázek 12: Struktura anotací v platformě MERLIN, celý korpus (<<MERLIN: anotace>>)

Zde můžete také získat přístup k **anotačním schématu**, na kterém jsou založeny všechny manuálně vytvořené anotace korpusu MERLIN, společně s příklady ze všech třech jazyků, dále glosář s méně běžnou terminologií a definice každého tagu. Navíc si zde můžete stáhnout souhrnnou **časosběrnou dokumentaci** o řešení obtížných situací, které při anotaci jednotlivých aspektů nastaly. Pokud narazíte na nějakou anotaci, která ve vás vyvolá pochybnosti, může vám být nápomocný právě tento dokument (dokument je také možné najít v sekci <<Dokumentace>>).

Pokud nechcete číst všechny definice tagů a stahovat kompletní anotační schéma, můžete se podívat na seznam všech anotačních tagů a příkladů k nim:

Seznam rysů žakovského jazyka s příklady

Gramatika | Pravopis | Srozumitelnost | Slovní zásoba | Koheze/Koherence | Sociolingvistická přiměřenost | Pragmatika

TAGY PRO GRAMATIKU	Příklad*
slovosled v hlavní větě	*[Vielleicht du könntest mir bei meine Wohnungssuche helfen.] *[Sollst du Wasser und Bikini mitbringen.]
slovosled ve vedlejší větě	*[wenn haben Sie Zeit,] dann bitte sagen Sie mir.
negace obecně	*Ich habe [nicht] Zeit.; *Er wird dort arbeiten [nein].
ČJ: dvojí negace	*[mám] žádný čas {nemám žádný čas}; *nikdo [volal] {nikdo nevolal}
valence slovesa: počet nezbytných doplňení	ČJ: *Petr vstává v 6 hodin. On nesnídá, protože [on] nemá hlad. NJ: *Er hat uns nicht gesagt, ob {er} kommen will.
shoda (podmět a přísudek)	*Jana [hast] gelesen, *Jana [sind] müde
zvrtné zájmeno	ČJ: *smála [si]; NJ: *er [entschuldigt], *Laura und Ferdinand reden [sich] IT: *[se] {si} lava ogni mattina
ČJ: přivlastňovací zvrtné zájmeno	*potřebuju [moji] knihu, *vidím [mého] otce
neexistující flexe (podstatná jména, přídavná jména a slovesa)	adjective: *ein [blaus] Himmel {blauer}; [teurer] {teurer}; [größen] {großen / größeren} noun: *das schöne [Hause], *[euche] [Fahrrade]

Obrázek 13: Seznam anotačních tagů s příklady (<<MERLIN: anotace>>)

3.5 <<Dokumentace>>



[Úvod](#) | [Vyhledávání](#) | [O projektu MERLIN](#) 

[cesky](#) [nemecky](#) [anglicky](#) [italsky](#)

MERLIN: v praxi	Dokumenty související s projektem MERLIN Dokumenty související s projektem MERLIN ke stažení.
MERLIN: výzkum	Korpus: koncipování testů a příprava dat Postup při transkripci (připravuje se ke stažení) Úplná zadání pro testy včetně popisu zadání, lze stáhnout v sekci » korpus MERLIN . » Hodnotící tabulka platformy MERLIN 
MERLIN: korpus	» Odborná zpráva  : Zpráva o reliabilitě a funkčnosti hodnotící škály MERLIN pro psané vzorky promluvy, autor: O. Bärenfänger
MERLIN: anotace	Anotace: Anotační schéma a jeho vytváření Anotační manuál (připravuje se ke stažení) » anotační schéma platformy MERLIN 
Dokumentace	Dokumentace týkající se dalších aspektů anotace (připravuje se ke stažení)
Ke stažení	Abstrakty z konferencí a publikace týmu MERLIN Katrin Wisniewski. <i>Die Validität der Skalen des Gemeinsamen europäischen Referenzrahmens für Sprachen. Eine empirische Untersuchung der Flüssigkeits- und Wortschatzskalen des GeRS am Beispiel des Italienischen und des Deutschen.</i> Language Testing and Evaluation vol.33, Frankfurt am Main 2014

Obrázek 14: <<Dokumentace>>

V této sekci jsou k nalezení všechny volně dostupné dokumenty týkající platformy MERLIN, tj.:

- projektové zprávy, např. o kvalitě hodnocení a použitelnosti platformy (výzkumy uživatelů)
- všechny použité testové úlohy
- tabulky použité pro hodnocení
- postupy a schémata použité pro transkripci a anotaci, např. anotační schéma
- publikace, prezentace apod. vytvořené týmem MERLIN

3.6 <<Stáhnout celý korpus>>

Stáhnout MERLIN korpus
MERLIN nabízí všechny texty obsažené v korpusu ke stažení zdarma.

(work in progress: to come soon)
Stáhněte si všechny žákovské texty včetně **metadat** (informace o autorovi textu, testu a hodnocení):

ČEŠTINA	soubory ve formátu PDF (zipované)	soubory ve formátu TXT (zipované)
NĚMČINA	soubory ve formátu PDF (zipované)	soubory ve formátu TXT (zipované)
ITALŠTINA	soubory ve formátu PDF (zipované)	soubory ve formátu TXT (zipované)

Stáhněte si všechny žákovské texty včetně **metadat a cílových hypotéz**:

ČEŠTINA	soubory ve formátu PDF (zipované)	soubory ve formátu TXT (zipované)
NĚMČINA	soubory ve formátu PDF (zipované)	soubory ve formátu TXT (zipované)
ITALŠTINA	soubory ve formátu PDF (zipované)	soubory ve formátu TXT (zipované)

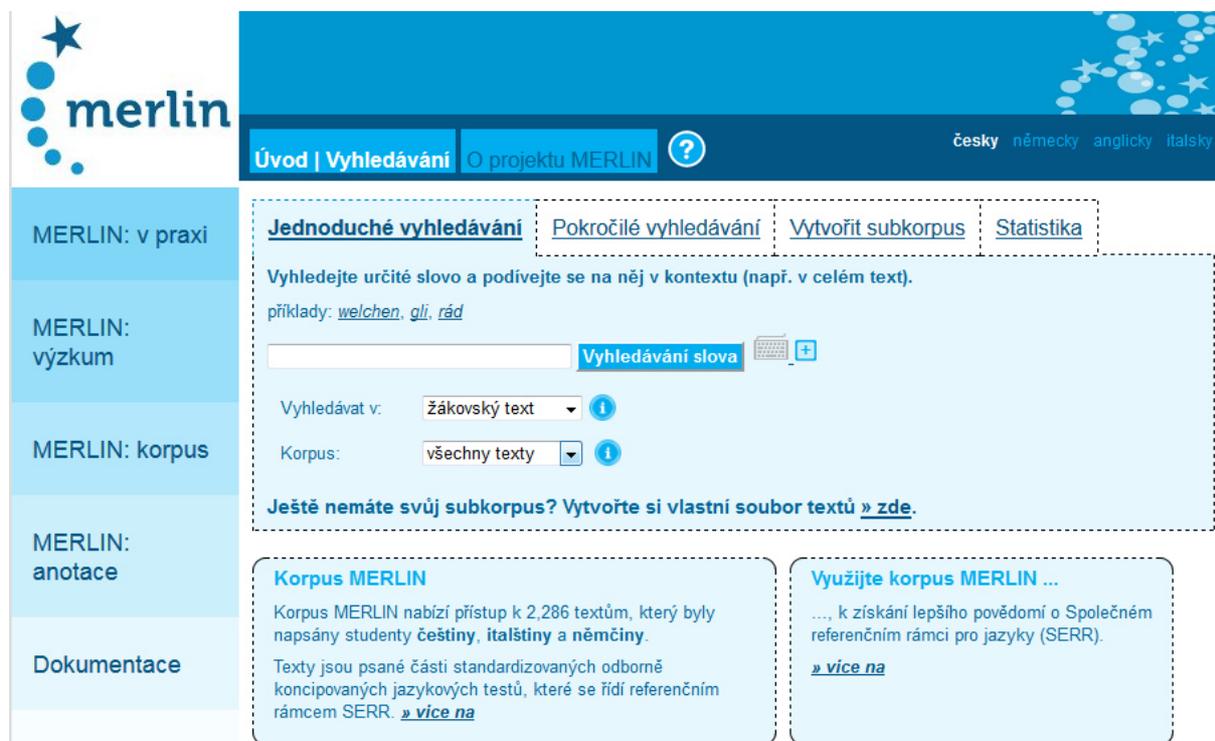
Jestli si chcete vytvořit vlastní soubor žákovských textů (subkorpus), použijte možnost **vytvořit subkorpus**. Odtud si můžete uložit vzorek nebo všechny texty subkorpusu jako jeden soubor ve formátu PDF nebo TXT s více stránkami.

Obrázek 15: <<Stáhnout celý korpus>>

V této sekci máte možnost okamžitého stahování celého korpusu MERLIN. Můžete si zde stáhnout zazipované soubory pro texty v češtině, italštině a němčině ve formátu .txt nebo .pdf. Máte přitom možnost zvolit, zdali si chcete stáhnout původní žákovské texty, obsahující metadata (např. mateřský jazyk nebo věk), anebo chcete také zahrnout cílové hypotézy.

Pokud by vás zajímal spíše víc specifický druh textů, jděte nejdříve do sekce <<Vytvořit subkorpus>>. V této sekci si můžete vytvořit subkorpus dle vašich potřeb, uložit si ho a vyhledávat v něm (<<Jednoduché / Pokročilé vyhledávání>>) nebo si nechat zobrazit statistické údaje (<<Statistika>>).

4 Vyhledávací funkce v platformě MERLIN



The screenshot shows the MERLIN search interface. On the left is a navigation menu with items: MERLIN: v praxi, MERLIN: výzkum, MERLIN: korpus, MERLIN: anotace, and Dokumentace. The main content area has a blue header with the MERLIN logo and navigation links: Úvod | Vyhledávání | O projektu MERLIN. There are also language options: česky, německy, anglicky, italsky. Below the header, there are four tabs: Jednoduché vyhledávání (selected), Pokročilé vyhledávání, Vytvořit subkorpus, and Statistika. The search form includes a text input field, a 'Vyhledávání slova' button, and dropdown menus for 'Vyhledávat v:' (set to 'žákovský text') and 'Korpus:' (set to 'všechny texty'). Below the form, there is a link: 'Ještě nemáte svůj subkorpus? Vytvořte si vlastní soubor textů » zde.' At the bottom, there are two informational boxes: 'Korpus MERLIN' and 'Využijte korpus MERLIN ...'.

Obrázek 16: Sekce vyhledávání

Do vyhledávacích možností patří:

- jednoduché vyhledávání
- pokročilé vyhledávání
- sekce “vytvořit subkorpus”
- statistická sekce

4.1 <<Vytvořit subkorpus>>

Uživatelům se doporučuje, aby začali používat možnosti vyhledávání definováním takového souboru textů, který je nejvíc zajímavá (<<Vytvořit subkorpus>>, viz Obrázek 17):

The screenshot shows the 'Vytvořit subkorpus' (Create subcorpus) configuration page in the MERLIN system. The page is part of a search interface, with tabs for 'Jednoduché vyhledávání', 'Pokročilé vyhledávání', 'Vytvořit subkorpus', and 'Statistika'. The 'Vytvořit subkorpus' tab is active. The main content area contains instructions: 'Vyberte text se kterým chcete pracovat. Získáte celý text a metadata (např. L1, věk, hodnocení úlohy). Použijte tento subkorpus pro další analýzu nebo ho můžete uložit.' Below this are several configuration options: 'Cílový jazyk' (Target language) set to 'všechny'; 'Úroveň testu podle SERR' (SERR test level) with a dropdown menu showing 'všechny', 'A1', 'A2', and 'B1'; 'Celkové hodnocení podle SERR' (Overall SERR rating) with a dropdown menu showing 'všechny', 'A1', 'A2', and 'A2+', and a link to 'Podrobná kritéria hodnocení'; and 'Testová úloha' (Test task) set to 'Nejprve pri'. There are also links for 'Vyberte texty podle informací o autorovi' and 'Vyberte texty podle výskytu slov a jazykových jevů'. At the bottom, there is a 'Název:' field and a 'Vytvořit subkorpus a zobrazit texty' button.

Obrázek 17: Rozhraní pro vyhledávání <<Vytvořit subkorpus>>

Zde je možné nastavit kritéria, podle kterých se může soubor žákovských textů, tzv. subkorpus, definovat dle potřeb uživatele. Mezi tato kritéria patří:

- cílový jazyk textu
- původní úroveň testu podle SERR
- adekvátní průměr hodnocení textů podle SERR (viz oddíl 2)
- jedna nebo více úrovní SERR v rámci jednotlivých hodnotících kritérií (gramatická správnost | rozsah slovní zásoby | ovládání slovní zásoby | sociolingvistická přiměřenost | koherence/koheze | pravopis)
- testová úloha

Dále je možné rozřídít texty podle údajů o autorovi:

- mateřský jazyk (L1)
- věk
- pohlaví

Texty mohou být zároveň vybírány podle obsahu maximálně tří rysů a/nebo slov v žákovském jazyce:

Vyberte texty podle výskytu slov a jazykových jevů

Vyhledávat v:

Slovo: ≥ počet výskytů

Slovo: ≥ počet výskytů

Slovo: ≥ počet výskytů

Jazykový jev: ≥ počet výskytů

Jazykový jev: ≥ počet výskytů

Jazykový jev: ≥ počet výskytů

Obrázek 18: Specifikace parametrů subkorpusu, rysy žákovského jazyka (<<Vytvořit subkorpus>>)

Uživatel pojmenuje subkorpus dle vlastní volby, a kliknutím na „Vytvořit subkorpus a zobrazit texty“ bude jeho subkorpus aktivní pro další vyhledávání po dobu následujících 24 hodin. Poté je nutné, aby uživatel subkorpus definoval znovu.

Výstup <<Vytvořit subkorpus>>

Kliknutím na “Vytvořit subkorpus a zobrazit texty” se zobrazí výstup, podobný tomu v grafice č. 19.

The screenshot shows the MERLIN search interface. The top navigation bar includes 'Úvod | Vyhledávání | O projektu MERLIN' and language options: 'česky', 'německy', 'anglicky', 'italsky'. The search filters section shows 'Váš subkorpus: lang=Czech; L1=all; CEFR level=all; overall rating=all'. Below this, there are buttons for 'použít subkorpus pro Jednoduché vyhledávání' and 'použít subkorpus pro Pokročilé vyhledávání'. The search results section shows 'Možnosti vyhledávání: Upravit předchozí dotaz', 'Vytvořit nový dotaz'. The results list shows 30 items (page 1 of 4), with 8 items per page. Each result includes a language level (e.g., 'Czech A2+ L1=German'), a button to 'Zobrazit žákovský text a TH', and a button for 'Informace o autorovi textu a hodnocení'. At the bottom, there is a section for 'Uložit texty' with a dropdown for 'Počet textů: 10', a checkbox for 'Včetně metadat', and a dropdown for 'Data: žákovský text' with an 'Uložit texty' button.

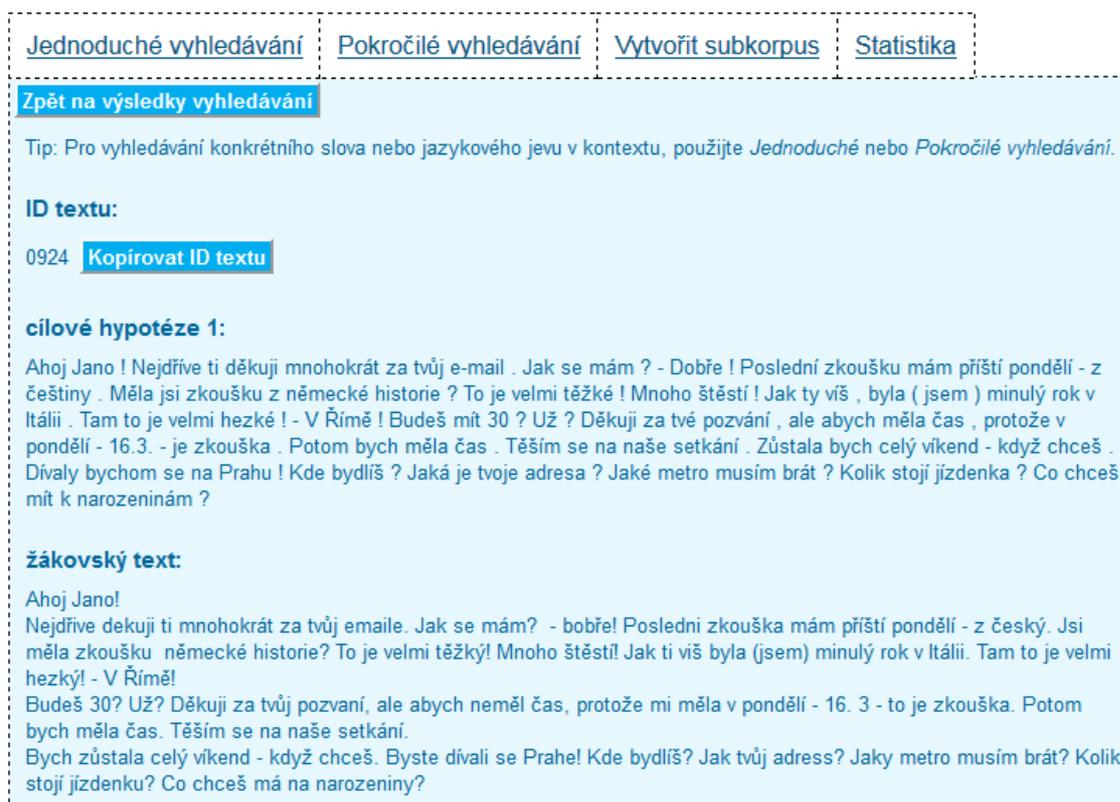
Obrázek 19: Příklad výstupu (<<Vytvořit subkorpus>>)

V tomto výstupu se zobrazí seznam všech textů, které vyhovují zadaným kritériím, která definují subkorpus. V našem příkladu uživatel shromáždil všechny česky psané texty, které se vztahují ke konkrétnímu testovému úkolu (není zobrazeno).

Na stránce s výstupem je specifikován počet nalezených textů („celkový počet výskytů: 30“). Z levého ID sloupce je vidět, o jaký cílový jazyk jde („čeština“), jaké je celkové hodnocení úrovně (různí se) a jaký je jazyk L1 (v uvedeném příkladu je to němčina).

Kliknutím na „stáhnout dokumenty“ můžete dokumenty **stáhnout** (anebo pouze určitý výběr z textů subkorpusu) buď s nebo bez metadat a s nebo bez cílových hypotéz v různých formátech.

Můžete rovněž kliknout na „**Zobrazit žákovský text a TH**“ a získat tak původní text a cílovou hypotézu 1 nebo 1 a 2 (příklad je uveden na obrázku č. 20).



[Jednoduché vyhledávání](#) [Pokročilé vyhledávání](#) [Vytvořit subkorpus](#) [Statistika](#)

[Zpět na výsledky vyhledávání](#)

Tip: Pro vyhledávání konkrétního slova nebo jazykového jevu v kontextu, použijte *Jednoduché* nebo *Pokročilé vyhledávání*.

ID textu:
0924 [Kopírovat ID textu](#)

cílové hypotéze 1:
Ahoj Jano ! Nejdříve ti děkuji mnohokrát za tvůj e-mail . Jak se mám ? - Dobře ! Poslední zkoušku mám příští pondělí - z češtiny . Měla jsi zkoušku z německé historie ? To je velmi těžké ! Mnoho štěstí ! Jak ty víš , byla (jsem) minulý rok v Itálii . Tam to je velmi hezké ! - V Římě ! Budeš mít 30 ? Už ? Děkuji za tvé pozvání , ale abych měla čas , protože v pondělí - 16.3. - je zkouška . Potom bych měla čas . Těším se na naše setkání . Zůstala bych celý víkend - když chceš . Dávaly bychom se na Prahu ! Kde bydlíš ? Jaká je tvoje adresa ? Jaké metro musím brát ? Kolik stojí jízdenka ? Co chceš mít k narozeninám ?

žákovský text:
Ahoj Jano!
Nejdříve dekuji ti mnohokrát za tvůj emaile. Jak se mám? - bobře! Posledni zkouška mám příští pondělí - z český. Jsi měla zkoušku německé historie? To je velmi těžký! Mnoho štěstí! Jak ti viš byla (jsem) minulý rok v Itálii. Tam to je velmi hezký! - V Římě!
Budeš 30? Už? Děkuji za tvůj pozvaní, ale abych neměl čas, protože mi měla v pondělí - 16. 3 - to je zkouška. Potom bych měla čas. Těším se na naše setkání.
Bych zůstala celý víkend - když chceš. Byste dávali se Prahe! Kde bydlíš? Jak tvůj adress? Jaky metro musím brát? Kolik stojí jízdenku? Co chceš má na narozeniny?

Obrázek 20: Příklad výstupu <<Zobrazit žákovský text a TH>> ve výstupu <<Definovat subkorpus>>

Kliknutím na „**Zobrazit informace o autorovi textu a hodnocení**“ se zobrazí metadata pro konkrétní text (Obrázek č. 21 znázorňuje metadata, která patří k textu v grafice č. 20):

Jednoduché vyhledávání | Pokročilé vyhledávání | Vytvořit subkorpus | Statistika

Zpět na výsledky vyhledávání

Tip: Pro vyhledávání konkrétního slova nebo jazykového jevu v kontextu, použijte *Jednoduché* nebo *Pokročilé vyhledávání*.

Obecně::

ID textu: 0924 [Kopírovat ID textu](#)

Cílový jazyk: Czech

Úroveň testu podle SERR: B1

Testová úloha: [answer to the email of Jana](#) [i](#)

Mateřský jazyk: German

Věk: not reported

Pohlaví: male

Hodnocení:

Celkové hodnocení podle SERR: A2+

Gramatická správnost: A2

Pravopis: B1

Slovní zásoba: rozsah: B1

Slovní zásoba: ovládnutí: A2

Koherence/koheze: A2

Sociolingvistická přiměřenost: A2

Obrázek 21: Příklad výstupu <<Zobrazit informace o autorovi textu a hodnocení>>ve výstupu <<Vytvořit subkorpus>>

V subkorpusu je dále možné vyhledávat v sekci jednoduchého nebo pokročilého vyhledávání nebo ve statistické sekci.

4.2 <<Jednoduché vyhledávání>>

V sekci <<Jednoduché vyhledávání>> je možné vyhledávat tvary slov:

The screenshot shows the MERLIN search interface. On the left is a navigation menu with items: MERLIN: v praxi, MERLIN: výzkum, MERLIN: korpus, MERLIN: anotace, Dokumentace, and Ke stažení. The main content area is titled 'Jednoduché vyhledávání' and includes a search bar, a dropdown for 'Vyhledávat v:' (set to 'žakovský text'), and a dropdown for 'Korpus:' (set to 'všechny texty'). Below the search area are two informational boxes: 'Korpus MERLIN' and 'Využijte korpus MERLIN ...'. The top navigation bar includes 'Úvod | Vyhledávání | O projektu MERLIN' and language options: 'česky | německy | anglicky | italsky'.

Obrázek 22: Rozhraní pro vyhledávání <<Jednoduché vyhledávání>>

Tento typ vyhledávání lze použít pro:

- žákovské texty a cílové hypotézy TH1 a TH2
- v celém korpusu MERLIN nebo v subkorpusu, který byl uživatelem definován (<<Vytvořit subkorpus>>)

Jednoduché vyhledávání neumožňuje pracovat s lemmaty, tj. se všemi slovními tvary daného slova, ale pouze s tím tvarem slova, který uživatel zadá. Takže např. po zadání slova „studovat“ do vyhledávání, jako v příkladu na obrázku č. 23, korpus nenabídne žádné výsledky typu „studuješ“ nebo „studuj“.

V případě potřeby můžete také použít *virtuální klávesnici*. Upozorňujeme, že platforma MERLIN pracuje s texty, které vznikly na základě omezeného počtu testových úloh a tím je zároveň omezen i rozsah obsažené slovní zásoby.

Výstup <<Jednoduchého vyhledávání>>

Výstup jednoduchého vyhledávání generuje slovo, které uživatel hledá, v jeho nejbližším kontextu (tzv. klíčové slovo v kontextu neboli KWIC):

The screenshot shows the MERLIN search interface. At the top left is the MERLIN logo. The navigation bar includes 'Úvod | Vyhledávání' and 'O projektu MERLIN'. Language options are 'česky', 'německy', 'anglicky', and 'italsky'. The main content area is titled 'Jednoduché vyhledávání' and shows search results for the word 'studovat'. It indicates that 64 occurrences were found in 43 documents. A pagination control shows 'Strana 1 z(e) 5 (15 počet výskytů na straně)'. Below this, there is a section 'Vyhledávané slovo v kontextu' with four examples of text snippets containing the word 'studovat'. To the right of each snippet is a blue button labeled 'Informace o autorovi textu a hodnocení'.

Obrázek 23: Příklad výstupu <<Jednoduché vyhledávání>>

Kliknutím na klíčové slovo v kontextu se zobrazí celý žákovský text. Kliknutím na <<Informace o autorovi textu a hodnocení>> se zobrazí metadata textu. Zároveň si můžete přímo stáhnout soubor s úlohou ve formátu .pdf a zkopírovat ID autora v případě, že chcete realizovat podrobnější vyhledávání:

The screenshot shows the MERLIN search interface. On the left is a navigation menu with options: MERLIN: v praxi, MERLIN: výzkum, MERLIN: korpus, MERLIN: anotace, Dokumentace, and Ke stažení. The top header contains the MERLIN logo, navigation links (Úvod | Vyhledávání, O projektu MERLIN), and language options (česky, německy, anglicky, italsky). The main content area is titled 'Zpět na výsledky vyhledávání' and includes a tip: 'Tip: Pro vyhledávání konkrétního slova nebo jazykového jevu v kontextu, použijte Jednoduché nebo Pokročilé vyhledávání.' Below this is a table of text characteristics:

Obecné::	
ID textu:	0924 Kopírovat ID textu
Cílový jazyk:	Czech
Úroveň testu podle SERR:	B1
Testová úloha:	answer to the email of Jana ⓘ
Mateřský jazyk:	German
Věk:	not reported
Pohlaví:	male
Hodnocení:	
Celkové hodnocení podle SERR:	A2+
Gramatická správnost:	A2
Pravopis:	B1
Slovní zásoba: rozsah:	B1
Slovní zásoba: ovládnání:	A2
Koherence/koheze:	A2
Sociolingvistická přiměřenost:	A2

Obrázek 24: Příklad výstupu <<Informace o autorovi textu a hodnocení>> ve výstupu <<Jednoduché vyhledávání>>

4.3 <<Pokročilé vyhledávání>>

Tato sekce vám umožňuje sofistikovanější způsob vyhledávání v korpusu MERLIN. Zde je možné zkombinovat vyhledávání lemat/slov s vyhledáváním různých anotací, které platforma MERLIN obsahuje.

Tuto funkci můžete využít pro vyhledávání:

- v žákovských textech nebo cílových hypotézách TH1 nebo TH2
- v celém korpusu MERLIN nebo subkorpusu, který byl uživatelem definován (<<Vytvořit subkorpus>>)

Obrázek 25: Rozhraní pro vyhledávání <<Pokročilé vyhledávání>>

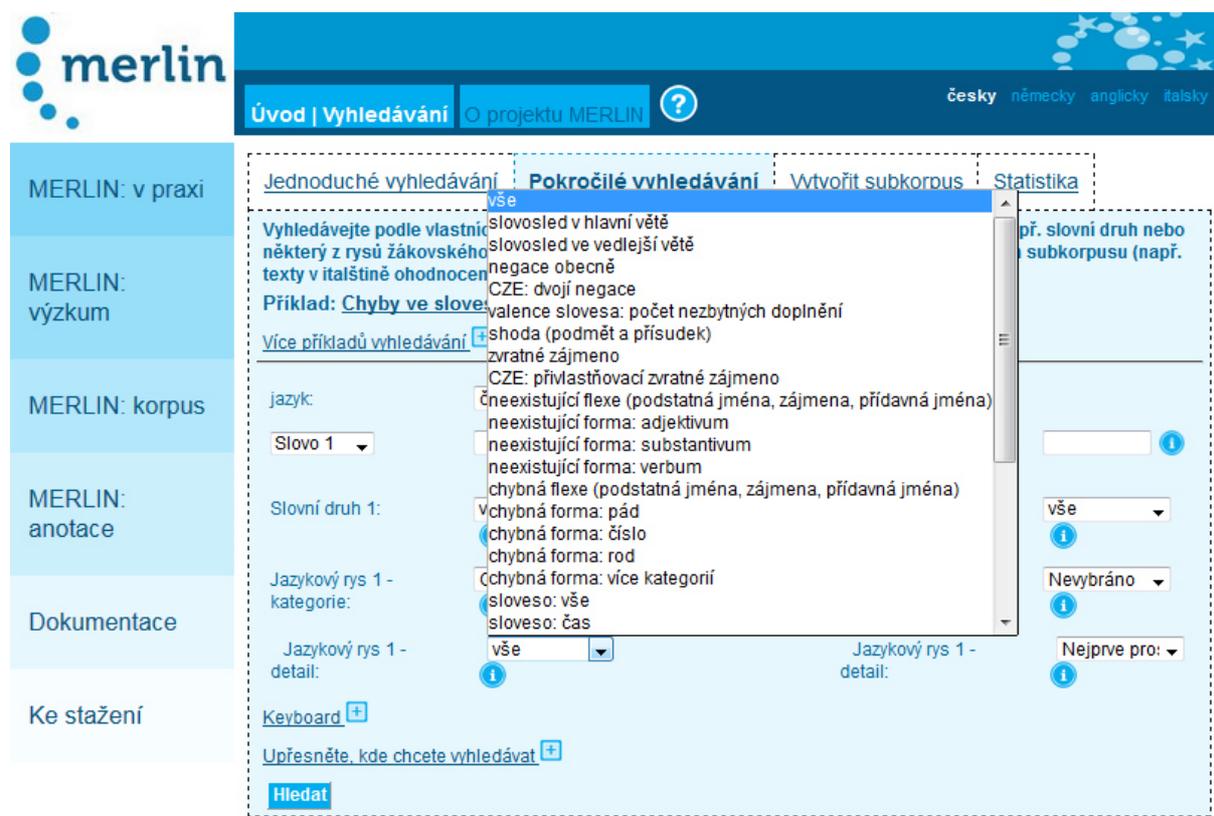
Dále se můžete rozhodnout, zdali chcete vyhledat 1 nebo 2 slova nebo lemmata, která:

- jsou přímo **sousedící** nebo se mezi nimi vyskytuje konkrétní počet slov
- Patří do určitého **slovního druhu**, který vás zajímá (na základě automatické anotace slovního druhu „POS“). Pokud vyberete cílový jazyk a kliknete na rozbalovací menu, zobrazí se nabídka zkratk slovních druhů a jejich popis (viz Obrázek č. 26).

Obrázek 26: Definování tagů slovních druhů , <<Pokročilé vyhledávání>>

Zde také získáte přístup k **manuálním anotacím**, které má platforma MERLIN k dispozici. Pro tento účel, vyberte

- nejdříve <<Jazykový rys 1 - kategorie>> pro specifikaci anotační kategorie, která vás zajímá, např. gramatika, slovní zásoba, pravopis
- Můžete také zúžit výběr pomocí <<Jazykový rys 1 - detail>>: v tomto případě získáte seznam všech **anotačních tagů**, které byly použity pro určitou anotační kategorii (Obrázek č. 27 zobrazuje příklad pro gramatiku). Pokud zde nevyberete žádný konkrétní tag, budou zobrazeny všechny gramatické tagy v případě, že vyberete kategorii „gramatika“ v <<Jazyková rys 1 - kategorie>>.



Obrázek 27: Definování <<Kategorie rysu>> v <<Pokročilém vyhledávání>> (detailně, gramatika)

Tento úkon je možné provést pro jedno slovo/lemma nebo pro kombinaci dvou slov/lemmat. Upozorňujeme, že vyhledávat rysy žakovského jazyka bez předešlé specifikace slova nebo lemmatu není možné.

Pokud potřebujete více informací o anotovaných rysech, můžete se podívat na seznam všech tagů a příkladů v sekci <<MERLIN: anotace>>; můžete se také dozvědět více o strukturní výstavbě anotací v platformě a stáhnout si její anotační schéma.

Výstup <<pokročilé vyhledávání>>

Sekce <<Pokročilé vyhledávání>> používá vyhledávač s otevřenou licencí „open source“ a vizualizační software ANNIS (www.annis-tools.org), kvůli kterému vypadá výstup v této

sekcí jinak než v ostatních výstupových sekcích platformy MERLIN. Následující ukázky stránek slouží jako průvodce výstupem.

Obrázek č. 28 znázorňuje zadávání slovesa „porodit“ v sekci <<Pokročilé vyhledávání>>.

The screenshot shows the MERLIN search interface. The top navigation bar includes 'Úvod | Vyhledávání | O projektu MERLIN' and language options: 'česky', 'německy', 'anglicky', 'italsky'. The main content area is titled 'Pokročilé vyhledávání' and contains the following search criteria:

- Language:** česky
- Lemma 1:** porodit
- Distance:** 0 vzdálenost slov
- Slovo 2:** (empty)
- Slovní druh 1:** vše
- Slovní druh 2:** vše
- Jazykový rys 1 - kategorie:** Nevybráno
- Jazykový rys 2 - kategorie:** Nevybráno
- Jazykový rys 1 - detail:** Nejprve prosí
- Jazykový rys 2 - detail:** Nejprve prosí

Additional options include 'Více příkladů vyhledávání', 'Keyboard', and 'Upřesněte, kde chcete vyhledávat'. A 'Hledat' button is located at the bottom of the search form.

Obrázek 28: Příklad vyhledávání lemma, <<Pokročilé vyhledávání>>

Výstup nabízí:

- (1) počet výskytů a počet dokumentů s výskyty (vlevo)
- (2) korpus, ve kterém jste vyhledávali (levá spodní část)
- (3) tlačítko, které skryje levou stranu výstupu, což je důležité proto, aby se žákovský text mohl zobrazit na celé obrazovce


- (4) možnost vrátit se zpět do sekce <<Pokročilé vyhledávání>> - výsledky hledání nebudou uloženy
- (5) odkaz s více informacemi (<<Vysvětlivky k tomuto výstupu>>), nové okno s informací o jednotlivých vrstvách anotace (viz kapitola č. 5 této publikace)
- (6) odkaz se seznamem všech zkratk s krátkým vysvětlením, zejména anotačních tagů, které byly ve výstupu použity (otevře se v novém okně) (<<Zkratky>>)
- (7) funkce nápovědy 

Zpět na Pokročilé vyhledávání | Vysvětlivky k tomuto výstupu | Zkratky

Help us to make ANNIS better! not logged in Login

tok_lemma="porodit"

Search More History

1 match in 1 document

Corpus List Search Options

Visible: All

Name	Texts	Tokens
2014-07-29-czech-gold-ea1-pilot	16	2 582
2014-07-29-czech-gold-pilot	436	78 899
2014-07-29-german-gold-ea1-pilot	355	51 413

Help/Examples Query Result

Base text Token Annotations

1 Path: MERLIN_Czech > 0720 left context: 5 right context: 5 (learner 213 - 218)

learner	porodit			moc	děti	.
TH1	porodit	bych	chtěla	moc	děti	.
TH1Diff		INS	INS			
TH2	porodit	bych	chtěla	moc	děti	.
TH2Diff		INS	INS			
EA_category		G_Verb_compl	G_Verb_main			
EA_category	H_Intelltxt					
EA_category	C_Coh_jump					
G_Verb_compl_type		o				
G_Verb_main_type			o			
H_Intelltxt_type	2					

transcript (grid) automatic (grid) dependencies (arcs) full text

Obrázek 29: Příklad výstupu vyhledávání lemma, <<Pokročilé vyhledávání>>

Když skryjete vertikální část výstupu vlevo, kde se zobrazují metainformace, kliknutím na



, bude přímé prohlížení žákovských textů snadnější:

[Zpět na Pokročilé vyhledávání](#) | [Vysvětlivky k tomuto výstupu](#) | [Zkratky](#)

[About ANNIS](#) not logged in [Login](#)

[Help/Examples](#) | Query Result

Base text | Token Annotations

1 / 1 | Displaying Results 1 - 1 of 1 | Result for: tok_lemma="porodit"

1 Path: MERLIN_Czech > 0720 (learner 213 - 218) | left context: 5 | right context: 5

TH/EA (grid)

learner	porodit			moc	děti	.
TH1	porodit	bych	chtěla	moc	děti	.
TH1Diff		INS	INS			
TH2	porodit	bych	chtěla	moc	děti	.
TH2Diff		INS	INS			
EA_category		G_Verb_compl	G_Verb_main			
EA_category	H_Intelltxt					
EA_category	C_Coh_jump					
G_Verb_compl_type		o				
G_Verb_main_type			o			
H_Intelltxt_type	2					

transcript (grid)
 automatic (grid)
 dependencies (arcs)
 full text

Obrázek 30: Příklad výstupu vyhledávání lemma, skrytá metainformace vlevo <<Pokročilé vyhledávání>>

Hledané lemma se zobrazí ve svém nejbližším kontextu, který můžete rozšířit až na 25 tokenů na každé straně:

left context: | right context:

0

5

10

15

20

25

Obrázek 31: Upravit rozsah kontextu pro lemma ve výstupu <<Pokročilého vyhledávání>>

Upozorňujeme, že ve výsledku vyhledávání můžete získat přístup k automatickým anotacím, celkovému zobrazení, transkriptu a závislostním obloukům. Na obrázku č. 32 je uveden příklad zobrazení automatické anotace:


 About ANNIS

[Help us to make ANNIS better!](#)

 not logged in [Login](#)

[Help/Examples](#)

[Base text](#)
[Token Annotations](#)

/ 1

 Result for: tok_lemma="porodit"

1 Path: MERLIN_Czech > 0720 (learner 213 - 218)

 left context: 5 right context: 5

TH/EA (grid)

learner	porodit			moc	děti	.
TH1	porodit	bych	chtěla	moc	děti	.
TH1Diff		INS	INS			
TH2	porodit	bych	chtěla	moc	děti	.
TH2Diff		INS	INS			
EA_category		G_Verb_compl	G_Verb_main			
EA_category	H_Intelltxt					
EA_category	C_Coh_jump					
G_Verb_compl_type		o				
G_Verb_main_type			o			
H_Intelltxt_type	2					

transcript (grid)

automatic (grid)

learner	porodit		moc	děti	.
tok_lemma	porodit		moc	dítě	.
tok_pos	Vf-----A----		NNFS4-----A----	NNFP2-----A----	Z:-----
tok_pos_bohnet	V		C	N	Z
sentence	Na diplomce chtěla bych pracovat v Stockholme X a porodit moc dětí.				
repetition			id:6 count:2		

dependencies (arcs)

full text

Obrázek 32: Zobrazit automatické anotace ve výstupu <<Pokročilého vyhledávání>>

Vysvětlení řádků v zobrazení výstupové tabulky naleznete po kliknutí na <<Vysvětlivky k tomuto výstupu>>.

V této sekci se nacházejí některé základní statistické údaje. Vyhledávat je můžete jak ve všech textech platformy MERLIN, tak v těch, které si zvolíte („subkorpus“, viz sekce <<Vytvořit subkorpus>>).

Obrázek 33: Rozhraní pro vyhledávání <<Statistika>>

Důležité upozornění pro interpretaci statistických údajů

Statistické údaje platformy MERLIN je nutné interpretovat s velkou opatrností. Údaje sice mohou naznačovat jisté tendence, ale nemohou být interpretovány až do takové míry, aby mohly sloužit k formulaci pravidel pro výuku jazyka. Je to z důvodů, z nichž některé zde uvádíme:

- 1) Korpus MERLIN je malý, takže není možné vyvodit jednoznačné obecné závěry.
- 2) Databáze anotací chyby EA1 a EA2 se od sebe liší; databáze EA2 je extrémně malá.
- 3) Kontrastivní analýzy by se měly provádět s opatrností. Ne vždy je smysluplné porovnávat statistické výsledky mezi jazyky navzájem, protože některé tagy jsou specifické jen pro některé z nich, např. chyby v užívání slovesného vidu, které jsou relevantní pouze v češtině a italštině, ale ne v němčině.
- 4) Statistické údaje úzce souvisí s množstvím úloh, které bylo pro platformu MERLIN použito (to platí hlavně pro slovní zásobu a oblasti EA2).

V sekci <<Statistika>> jsou k dispozici **čtyři typy údajů**. Při jakékoliv volbě můžete vybrat několik anotovaných rysů zároveň podržením klávesy CTRL:

(1) četnost výskytu anotovaných rysů

Absolutní četnost anotovaných rysů ve vašem subkorpusu. Můžete si vybrat jednu nebo více kategorií rysů, např. gramatika nebo pravopis. Poté můžete vyhledat buď počet všech anotací v této kategorii/-ích, např. všechny gramatické anotace v německých textech, nebo specifikovat, které anotace vás zajímají na konkrétnější úrovni, např. všechny morfologické chyby v německých textech:

The screenshot shows the 'Statistika' tab in a web application. It features a search interface with the following elements:

- Navigation tabs: 'Jednoduché vyhledávání', 'Pokročilé vyhledávání', 'Vytvořit subkorpus', and 'Statistika' (active).
- Text: 'Zobrazit frekvenci výskytu jevů ve vybraných souborech textů (subkorpusech)'. Below it, a link 'Prosím mějte na paměti: [+]'.
- Form fields: 'Korpus:' with a dropdown menu (options: 'all Czech texts', 'all German texts', 'all Italian texts'), and 'Text (ID textu):' with an empty input field.
- Text: 'Ještě nemáte svůj subkorpus? Vytvořte si vlastní soubor textů » zde.'
- Section: 'Frekvence anotovaných jevů' with a minus sign and an information icon.
- Category list: 'GRAMMATIKA' with a minus sign. A scrollable list contains: 'všechny', 'slovosled v hlavní větě', 'slovosled ve vedlejší větě', 'negace obecně', and 'CZE: dvojitá negace'.
- Other categories: 'PRAVOPIS [+]', 'SROZUMITELNOST [+]', 'SLOVNÍ ZÁSoba [+]', 'KOHERENCE/KOHEZE [+]', 'SOCIOLINGVISTICKÁ PŘÍMĚŘENOST [+]', and 'PRAGMATIKA [+]'.
- Additional options: 'Relativní frekvence anotovaných jevů [+]', 'Bezchybná žakovská produkce [+]', and 'Komplexita textu (dostupné pouze pro němčinu) [+]'.
- Button: 'Zobrazit'.

Obrázek 34: Příklad výpočtu absolutní frekvence (všechny gramatické chyby, němčina), <<Statistika>>

Následující **výstup** podává přehled celkového počtu vyhledávaných rysů/-ů:

The screenshot shows the 'Statistika' tab displaying search results. It includes:

- Navigation tabs: 'Jednoduché vyhledávání', 'Pokročilé vyhledávání', 'Vytvořit subkorpus', and 'Statistika' (active).
- Text: 'Upravit předchozí dotaz'.
- Text: 'Frekvence rysů žakovského jazyka ve vybraných textech:'.
- Table:

Corpus	German
Počet textů	1034
Frekvence anotovaných jevů	
GRAMMAR	4749

Obrázek 35: Výstup příkladu výpočtu absolutní frekvence (všechny gramatické chyby, němčina), <<Statistika>>

Kliknutím na číslo počtu výskytů (v našem příkladu to je 4749) budete přesměrováni do sekce <<Pokročilé vyhledávání>>, kde získáte přímý přístup ke všem anotacím, které jsou pro vaše vyhledávání relevantní:

The screenshot shows the ANNIS interface with the following elements:

- Navigation tabs: [Back to advanced search](#), [About this search output](#), [Abbreviations](#)
- Search bar: [About ANNIS](#), [Help us to make ANNIS better!](#), [not logged in](#), [Login](#)
- Help/Examples, Query Result
- Base text, Token Annotations
- Navigation: [1](#) / 473, [Displaying Results 1 - 10 of 4722](#), [Result for: scheme=/G_*/](#)
- Result 1: Path: 2014-07-29-german-gold-pilot > 1023_0001416 (learner 68 - 78) left context: 5 right context: 5
- Annotations table:

Annotation	1	2	3	4	5	6	7	8	9	10	11	12
TH1/EA1 (grid)												
learner	Ich	bin	flexibel	und	für	neuen	Aufgaben	offen	.	Takt	,	.
ZH1	Ich	bin	flexibel	und	für	neue	Aufgaben	offen	.	Takt	,	.
ZH1Diff						CHA		CHA				
schema						G_Morphol_Wrong		O_Graph				
g_morphol_wrong_type						ambig						
o_graph_graphgen_act_type								o				
o_graph_type								graphgen				

Result 2: Path: 2014-07-29-german-gold-pilot > 1023_0001416 (learner 114 - 126) left context: 5 right context: 5

Obrázek 36: Všechny výsledky příkladového výpočtu frekvence (Obrázek č. 34-35) ve výstupu <<Pokročilé vyhledávání>>

(2) relativní četnost anotovaných rysů

Relativní počet anotovaných rysů ve vašem subkorpusu v jedné větě nebo tokenu. Jelikož korpus obsahuje různě dlouhé texty, je užitečné používat normalizovanou četnost, tj. počet výskytů jednoho nebo více jevů s ohledem na standardizovaný subjekt, např. jako v našem případě, věty a tokeny. Pomocí této funkce byste například mohli vypočítat průměrný počet gramatických chyb v žákovských textech, které byly ohodnoceny úrovní B1, oproti textům na úrovni B2 (nejdříve ale definujte subkorpusy).

Postup zobrazování relativní četnosti je analogický k výše uvedenému postupu zobrazování absolutní četnosti. Z výstupu v sekci <<Statistika>> můžete kliknutím na výsledek vyhledávání rovněž získat přístup k příkladům z korpusu ve výstupu <<Pokročilé vyhledávání>>.

(3) bezchybná žákovská produkce

Tyto údaje vycházejí z pozitivní perspektivy vůči žákovskému jazyku, což znamená, že je zobrazeno určité procento jazyka, které **neobsahuje** žádné rysy s anotací chyby. Procento bezchybných vět nebo tokenů lze vypočítat pomocí celkového počtu vět nebo tokenů. Na

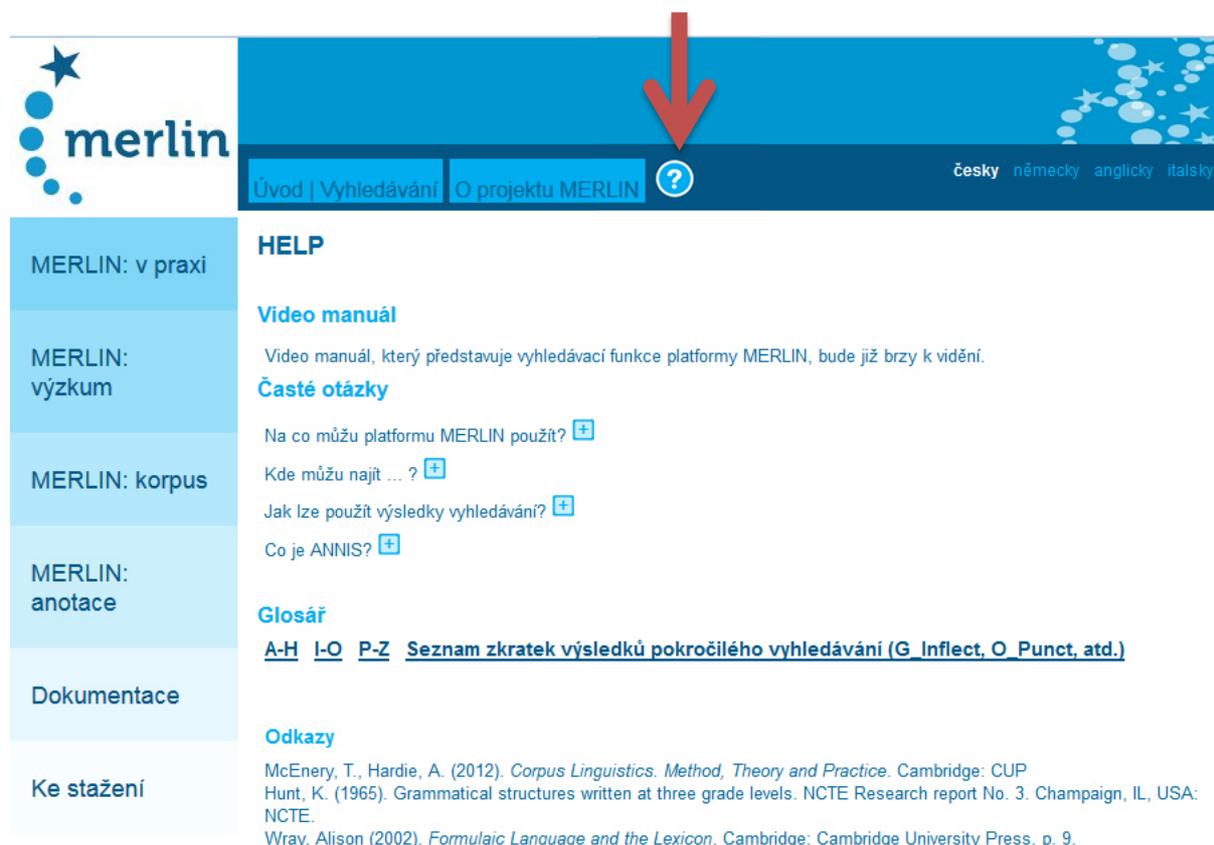
detailnější úrovni můžete také zjistit procento jazyka bez gramatických anebo morfologických chyb nebo chyb v psaní velkých písmen.

(4) komplexita (pouze pro němčinu)

Do statistické sekce bylo pro němčinu možné zahrnout také automaticky počítané údaje komplexity morfologických, lexikálních a syntaktických prvků. Pro italštinu a češtinu nebyly bohužel k dispozici potřebné technické prerekvizity.

Komplexita textu je důležitým aspektem (žákovského) jazyka, který je úzce spojen s úrovní ovládnutí jazyka. Mnoha výzkumy byla komplexita poměrně jasně odlišena od přesnosti či plynosti (jde o tzv. výzkumy CAF, viz bibliografie). Údaje, ke kterým mají uživatelé platformy MERLIN přístup, pocházejí z výzkumů z oblasti hodnocení komplexity a čitelnosti druhého jazyka. Poprvé byly použity Hanckem (2013) a Hanckem & Meurerem (2013) s ohledem na data platformy MERLIN (<<*Dokumentace*>>).

5 Nápověda



The screenshot shows the MERLIN website interface. At the top left is the MERLIN logo. The top navigation bar contains links for 'Úvod', 'Vyhledávání', 'O projektu MERLIN', and a help icon (a question mark in a circle). A red arrow points to this help icon. Below the navigation bar is a sidebar with categories: 'MERLIN: v praxi', 'MERLIN: výzkum', 'MERLIN: korpus', 'MERLIN: anotace', 'Dokumentace', and 'Ke stažení'. The main content area is titled 'HELP' and contains sections for 'Video manuál', 'Časté otázky', 'Glosář', and 'Odkazy'. The 'Časté otázky' section lists several questions with plus icons for expansion. The 'Glosář' section lists abbreviations A-H, I-O, P-Z and a link to a list of abbreviations. The 'Odkazy' section lists three references.

Obrázek 37: Nápověda

Rozhraní MERLIN nabízí mnoho dokumentů či funkcí, vám pomohou v lepší orientaci a pochopení všech funkcí platformy. Do sekce nápovědy () se můžete dostat z jakéhokoli místa rozhraní.

5.1 Uživatelský manuál

Tento *Uživatelský manuál* je na rozhraní platformy MERLIN k dispozici na stránce <<Nápověda>> v němčině, italštině, angličtině a češtině a bude pravidelně aktualizován.

5.2 Seznámení se s platformou MERLIN: video instruktáž

Můžete zhlédnout dva videozáznamy (asi 30 minut) o projektu MERLIN (anglicky hovoří Katrin Wisniewski). Jedno video je určeno vyučujícím a druhé se více orientuje na jazykové testery a autory učebnic. Prezentace byly zaznamenány během dvou workshopu, které se konaly v prosinci 2014 v Linci.

5.3 Navigace rozhraním MERLIN: manuál s ukázkami

Manuál s ukázkami jednotlivých sekcí platformy, který je k dispozici v angličtině a němčině), vás provede jednotlivými funkcemi rozhraní platformy MERLIN.

Je rozdělen do dvou bloků: první, základní část (*sekce <<Vytvořit subkorpus>>, <<Jednoduché vyhledávání>>*) vás naučí první kroky při používání rozhraní, které spočívají v sestavení a/nebo uložení souboru textů/úloh, které vás zajímají a vyhledávání slov v nich. V druhé části je vysvětleno, jak funguje *<<Pokročilé vyhledávání>>* lemmat nebo anotací a zároveň krátce popisuje sekci *<<Statistika>>*.

5.4 Jak používat platformu MERLIN – příklady užití

V prosinci 2014 zorganizoval tým projektu MERLIN několik workshopů v rakouském Linci, kde představil rozhraní MERLIN a možnosti jeho využití učiteli jazyků, jazykovými testery a školiteli. Tyto materiály jsou volně dostupné ke stažení v sekci *<<MERLIN: v praxi>>* v češtině, italštině i němčině. Modelové scénáře byly vytvořeny v prosinci 2014, když ještě nebyly plně v provozu všechny funkce platformy projektu MERLIN. Z toho důvodu se mohou některé screenshoty a ukázky odlišovat od současné podoby platformy na webu.

5.5 Porozumění výstupu <<Pokročilé vyhledávání>>

Jak už bylo zmíněno výše (kapitola 4.3), používá sekce *<<Pokročilé vyhledávání>>* vyhledávač s otevřenou licencí „open source“ a vizualizační software ANNIS. Kromě využívání původních funkcí ANNISu zakomponoval tým projektu MERLIN do výstupů vyhledávání ještě několik funkcí nápovědy, díky nimž je používání tohoto typu vyhledávání srozumitelnější.

The screenshot shows the MERLIN platform interface. At the top, there is a navigation bar with the MERLIN logo and a menu with items: 'Úvod | Vyhledávání', 'projektu MERLIN', and a help icon. Below this is a sub-menu with 'Zpět na Pokročilé vyhledávání', 'Vysvětlivky k tomuto výstupu', and 'Zkratky'. The main content area shows a search result for 'unavený' with a grid of TH/EA annotations. A red arrow points to the 'Úvod | Vyhledávání' menu item.

learner	.	takže	teď	jsem	jenom	unavená	.	Měj	se	hezky	.
TH1	.	takže	teď	jsem	jenom	unavený	.	Měj	se	hezky	.
TH1Diff						CHA				CHA	
TH2	.	takže	teď	jsem	jenom	unavený	.	Měj	se	hezky	.
TH2Diff						CHA				CHA	
EA_category	H_Intellts							V_FS		G_Pos	
EA_category						G_Morphol_Wrong		V_FS			

Obrázek 38: Nápověda ve výstupu <<Pokročilé vyhledávání>>

Kliknutím na <<Vysvětlivky k tomuto výstupu>> se otevře nové okno, aby bylo umožněno pokračovat v analýze výstupu vyhledávání. Všechny vrstvy tabulky výstupu pokročilého vyhledávání jsou vysvětleny na následujícím obrázku č. 39.

merlin

Úvod | Vyhledávání | O projektu MERLIN ?

česky německy anglicky italsky

Vysvětlení výsledků vyhledávání

Upozornění: text je zobrazen v novém okně (na nové záložce vašeho prohlížeče).

Výsledky vašeho vyhledávání se zobrazí v aplikaci ANNIS. ANNIS je "open-source" software (v podstatě jde o vyhledávací a vizualizační systém), který zobrazuje anotace na několika úrovních jako ty z korpusu MERLIN. Umožňuje uživateli s korpusu pracovat s celým souborem různých typů anotací v rámci programu MERLIN: s cílovými hypotézami (TH1, TH2), anotacemi rysů žákovského jazyka a automaticky přiřazenými anotacemi (např. slovní druhy, věty atd.).

Ve vyhledávacím poli na levé straně (viz [1] na obrázku níže) vidíte průběh vyhledávání prostřednictvím vyhledávacího rozhraní MERLIN (pokročilé vyhledávání) přeložené do vyhledávacího jazyka aplikace ANNIS. Chcete-li změnit parametry vyhledávání, můžete si znovu otevřít rozhraní MERLIN („zpět na pokročilé vyhledávání“) nebo jej upravit pomocí vyhledávacího jazyka ANNIS.

» Více informací o programu ANNIS a vyhledávacím jazyce naleznete na » úvodní stránce aplikace ANNIS

Základní informace o výsledcích vašeho vyhledávání v programu ANNIS

47 matches in 37 documents

learner	eine	Bade	Zimmer	ein	Schlaf	Zimmer	eine	Küche	Un
ZH1	ein	Badezimmer	ein	Schlafzimmer	eine	Küche	haben	unc	
ZHDI#	CHA	MERGE	INS	MERGE	INS		DEL	CH	
ZHISpec							1		
schema	G_Morphol_Wrong	O_Wordbd		O_Punct					
schema	O_Punct								
schema	G_Verb_main			O_Wordbd					
g_morphol_wrong_type	gend								
g_verb_main_type	o								
o_punct_type	o								
o_wordbd_type	split			split					

Obrázek 39: Funkce nápovědy <<Vysvětlivky k tomuto výstupu >> ve výstupu <<Pokročilého vyhledávání>>

Další možností nápovědy je otevřít soubor se seznamem použitých zkratk <<Zkratky>> (Obrázek č. 40). Anotační tagy s krátkými popisy se zobrazí v novém okně.

merlin

Úvod | Vyhledávání | O projektu MERLIN ?

česky německy anglicky italsky

Slovníček zkratk (anotační tagy)

Upozornění: text je zobrazen v novém okně (na nové záložce vašeho prohlížeče).

Prvky schématu (scheme)

Úroveň schéma obsahuje informace o anotovaných rysech žákovského jazyka. Niže uvedená schémata postihují všechny zkratky a jejich význam. Více příkladů jazykových rysů viz » MERLIN anotace. Podrobné informace o různých typech anotací, popisech tagů a rozsahu tagů, viz » anotační schéma

G_ Gramatika		
G_Agr		shoda (podmět a přísudek)
G_Art		člen
G_Clit		ITA: příklonka
G_Conj		spojka
G_Infect_inexist		neexistující flexe (podstatná jména, přídavná jména, slovesa)
G_Morphol_wrong		chybná flexe (podstatná jména, zájmena, přídavná jména)

Obrázek 40: Seznam zkratk používaných ve výstupu <<Pokročilého vyhledávání>>

5.6 Glossář

Glosář, který je k nalezení v sekci nápovědy () obsahuje termíny užívané v rozhraní platformy MERLIN.

5.7 Časté otázky

Tato část shrnuje otázky uživatelů. Najdete zde např. informace o tom, k jakým účelům lze platformu MERLIN používat, jak pracovat s výsledky vyhledávání nebo je zde také vysvětlen výstup <<Pokročilého vyhledávání>>. Tato část bude pravidelně aktualizována.

5.8 Kontakt

Pokud máte jakékoliv otázky nebo podněty, prosíme, neváhejte kontaktovat tým projektu MERLIN (e-mail: info@merlin-platform.eu). Rádi vám pomůžeme.

Literatura

- [ALTE 2001] = ALTE Working Group on the Code of Practice: *Principles of Good Practice for ALTE Examinations*. Revised Draft. http://www.testdaf.de/institut/pdf/ALTE/ALTE_good_practice.pdf, December 2014.
- [Consiglio d'Europa 2004a] = Trim, J./North, B./Coste, D.: *Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione*. La Nuova Italia: Oxford.- A cura del Consiglio d'Europa.
- [Council of Europe 1975] = Van Ek, J. A.: *The Threshold Level in a European unit/credit system for modern language learning by adults*. Strasbourg: Council of Europe.
- [Council of Europe 1994a] = North, B.: *Scales of language proficiency: a survey of some existing systems*. Strasbourg: Council of Europe, CC-Lang (94) 24.
- [Council of Europe 1994b [1981]] = Galli de' Paratesi, N.: *Livello Soglia per l'insegnamento dell'italiano come lingua straniera*. Strasbourg: Edizioni del Consiglio d'Europa.
- [Council of Europe 1999 [1980]] = Baldegger, M./Müller, M./Schneider, G. (1999): *Kontaktschwelle Deutsch als Fremdsprache*. 4. Auflage. Berlin u.a.: Langenscheidt.- ed. by Council of Europe.
- [Council of Europe 2001a] = Trim, J./North, B./Coste, D.: *Common European Framework of Reference for Languages: Learning, teaching, assessment*. -Edited by the Council of Europe. Online-Dokument: www.coe.int/lang, December 2014.
- [Council of Europe 2001b] = Trim, J./North, B./Coste, D.: *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. Berlin u.a.: Langenscheidt.- Herausgegeben vom Europarat, Online-Dokument: <http://www.goethe.de/z/50/commeuro/i7.htm>, December 2014.
- [Europarat 2004] = Takala, S./Kaftandjieva, F./Verhelst, N./Banerjee, J./Eckes, T./van der Schoot, F.: *Reference Supplement to the Preliminary Pilot Version of the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*.- Edited by the Council of Europe, Online-Dokument: www.coe.int/lang, December 2014.
- [Europarat 2009 [2003]] = North, B./Figueras, N./Takala, S./Van Avermaet, P./Verhelst, N.: *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Manual. Preliminary Pilot Version*.- Edited by the Council of Europe, Online-Dokument: www.coe.int/lang, December 2014.
- Abel, A. / Wisniewski, K. / Nicolas, L. / Boyd, A. / Hana, J. / Meurers, D. (2014): A Trilingual Learner Corpus illustrating European Reference Levels. In: *Ricognizioni – Rivista di Lingue, Letterature e Culture Moderne* 2 (1), 111-126. (<http://www.ojs.unito.it/index.php/ricognizioni>).
- Abel, A. / Glaznieks, A. / Nicolas, L. / Stemle, E. (2014): KoKo: an L1 Learner Corpus for German. In: Calzolari, N. / Choukri, K. / Declerck, Th. / Loftsson, H./ Maegaard, B. / Mariani, J. / Moreno, A. / Odijk, J. / Piperidis, St. (eds.): *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, 26–31 May 2014. Reykjavik: European Language Resources Association (ELRA). 2414-2421. (<http://www.lrec-conf.org/proceedings/lrec2014/index.html>).
- Abel, A. / Wisniewski, K. (2012): Sprechaktrealisierungen in der L2 und der GERS: Ein- und Aussichten für Sprachwissenschaft und Didaktik. In: Di Meola, Claudio / Hornung, Antonie / Rega, Lorenza (Hrsgg.): *Perspektiven Vier. Akten der 4. Tagung Deutsche Sprachwissenschaft in Italien*. Rom, 4.-6.Februar 2010. Frankfurt a.M. 311-325.
- Abel, A. (2010): Sprachtests und soziale Implikationen. In: *Deutsch als Fremdsprache*, 4/2010. 202-209.

- AERA/APA/NCME (1999): *Standards for educational and psychological testing*. Washington: AERA.
- Alderson, J.C. (1991): Bands and scores. In: Alderson, J.C./North, B. (eds.): *Language testing in the 1990s*. London: British Council/Macmillan, 71-86.
- Aguado, K. (2004): Evaluation fremdsprachlicher Wortschatzkompetenz: Funktionen, Prinzipien, Charakteristika, Desiderate. In: Tschirner (Hrsg.) 231- 250.
- Aijmer, K. (1996), *Conversational routines in English. Convention and creativity*. London/NY: Longman.
- Alderson, J. C./Figueras, N./Kuijper, H./Nold, G./Takala, S./Tardieu, C. (2006): Analysing Tests of Reading and Listening in Relation to the Common European Framework of Reference: The Experience of the Dutch CEFR Construct Project. In: *Language Assessment Quarterly* 3(1), 3-30.
- Alderson, J.C. (2007): The CEFR and the need for more research. In: *The Modern Language Journal* 91, 658-662.
- Al-Gahtani, S./Roever, C. (2012): Proficiency and Sequential Organization of L2 REQUESTs. In: Applied Linguistics 33/1, 42 –65.
- Al-Jarf, R. Spelling error corpora in EFL. In the proceedings of the International Conference on Multi Development and Application of Language and Linguistics, National Cheng Kung University, May 5-16 2009, Tainan City, Taiwan, 2009.
- Arnaud, P. J. L. (1984): The lexical richness of L2 written productionos and the validity of vocabulary tests: In: Culhane, T./Klein-Braley, C./Stevenson, D. K. (eds.): *Practice and Problems in Language*
- Arnaud, P.J.L. (1984): The lexical richness of L2 written productions and the validity of vocabulary tests. In: Culhane, T./Klein-Braley, C./Stevenson, D.K. (eds.): *Practice and Problems in Language Testing*. Essex: Department of Language and Linguistics, University of Essex, 14-28.
- Arras, U. (2010): Subjektive Theorien als Faktor bei der Beurteilung fremdsprachlicher Kompetenzen. In: Berndt, A./Kleppin, K. (eds.): *Sprachlehrforschung: Theorie und Empirie - Festschrift für Rüdiger Grotjahn*. Frankfurt: Lang, 169-179.
- Bachman, L.F. (1990): *Fundamental Considerations in Language Testing*. New York: OUP.
- Bachman, L.F. (2004): *Statistical analyses for language assessment*. Cambridge: CUP 2004.
- Bachman, L.F./Palmer, A. (1996): *Language Testing in Practice*. New York: OUP.
- Bachman, L.F./Palmer, A. (2010): *Language Testing in Practice. Developing Language Assessment and Justifying their Use in the Real World*. Oxford: OUP.
- Bachmann, T. (2002): *Kohäsion und Kohärenz: Indikatoren für Schreibentwicklung: Zum Aufbau kohärenzstiftender Strukturen in instruktiven Texten von Kindern und Jugendlichen*. Innsbruck: Studienverlag.
- Baker, P. (2010): *Sociolinguistics and Corpus Linguistics*. Edingburgh: EUP.
- Bardovi-Harlig, K. & Bofman, T. (1989) Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11 (1), 17-34.
- Bardovi-Harlig, K. (2009): Conventional Expressions as a Pragmalinguistic Resource: Recognition and Productions of Conventional Expressions in L2 Pragmatics. In: *Language Learning* 59 (4), 755-795.
- Bardovi-Harlig, K. (2013): Developing L2 Pragmatics. In: *Language Learning* 63 (1): Suppl. 1, 66-86.
- Barron, A. (2003), *Acquisition in Interlanguage Pragmatics. Learning How to do Things with Words in a Study Abroad Context*. Amsterdam/Philadelphia: Benjamins.

- Bausch, K.-R./Christ, H./Königs, F.G./Krumm, H.-J. (eds.) (2003): *Der Gemeinsame Europäische Referenzrahmen für Sprachen in der Diskussion. Arbeitspapiere der 15. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr.
- Bayley, R. (2007): Second language acquisition: a variationist perspective. In: Bayley, R. /Lucas, C. (eds.) (2007): *Sociolinguistic Variation : Theories, Methods, and Applications*. Cambridge University Press, 133-144.
- Bayley, R. / Regan, V. (eds) (2004). *The acquisition of sociolinguistic competence*. Special Issue of the *Journal of Sociolinguistics*, 8 (3).
- Beebe, L. (1988). Five sociolinguistic approaches to Second Language Acquisition. In: L.Beebe (ed.), *Issues in Second Language Acquisition: Multiple Perspectives*. Cambridge, MA: Newbury House, pp. 43–75.
- Bestgen, Y./Granger, S. (2011): Categorising spelling errors to assess L2 writing. In: *International Journal of Continuing Engineering Education and Life Long Learning*, 21 (2), 235-252.
- Biber, D./Finegan, E. (eds.) (1994): *Sociolinguistic perspectives on register*. New York: OUP.
- Blum-Kulka, S. (1987), Indirectness and politeness in requests: Same or different? *Journal of Pragmatics* 11, 1, 131-46.
- Blum-Kulka, S. (1991), Interlanguage pragmatics: The case of requests. In Phillipson, R./Kellerman, E./Selinker, L./Sharwood Smith, M./Swain, M. (eds.) (1991), *Foreign/second language pedagogy research: A*
- Blum-Kulka, S./House, J./Kasper, G. (eds.) (1989), *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex.
- Blum-Kulka, S./Olshtain, E. (1984), Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics* 5, 3, 196-213.
- Bond, T. G./Fox, C. M. (2007): *Applying the Rasch model: Fundamental measurement in human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Botley, S. and Dillah, D. (2007) Investigating spelling errors in a Malaysian learner corpus. *Malaysian Journal of ELT Research*, Vol. 3, pp.74-93.
- Bredel, U. (2010) (ed.) : *Schriftsystem und Schrifterwerb: linguistisch – didaktisch – empirisch*. Berlin: de Gruyter.
- Bulté, B./Housen, A. (2012): Defining and operationalising L2 complexity. In: Housen, A./Kuiken, F./Vedder, I. (eds.): *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: Benjamins, 21-46.
- Burger, H. (2007): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. (3. Aufl.).Berlin: Erich Schmidt Verlag.
- Carlsen, C. (2010): Discourse connectives across CEFR levels: A corpus-based study. In: Bartning, I./Martin, M./Vedder, I. (eds.): *Communicative Proficiency and Linguistic Development: intersections between SLA and language testing research* (Eurosla). 191-210. purl.org/net/Carlsen-10.pdf.
- Carlsen, C. (2010); Linking a learner corpus to the Common European Framework of Reference. Manuscript submitted for publication.
- Carlsen, C. (ed.) 2013. *Norsk Profil. Det felles europeiske rammeverket spesifisert for norsk. Et første steg*. Oslo: Novus.
- Casanave, C. (1994) Language development in students' journals. *Journal of Second Language Writing*. 3, 179-201.
- Castro, C. D. (2004): Cohesion and the social construction of meaning in the essays of Filipino college students' writings in L2 English, in: *Asia Pacific Education Review*, 5, 215-225.

- Chastain, K. (1990) Characteristics of graded and undergraded compositions. *Modern Language Journal*, 74, 10-14.
- Chen, M., Zechner, K. (2011) Computing and Evaluating Syntactic Complexity Features for Automated Scoring of Spontaneous Non-Native Speech. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oregon, June 19-24, 2011. ACL, pp. 722–731.
- Chiang, St. (2003): The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning, in: *System*, Vol.31(4), 471-484.
- Cho, Y. (2005) :Grammatik und Höflichkeit im Sprachvergleich. Direktive Handlungsspiele des Bittens, Aufforderns und Anweisens im Deutschen und Koreanischen. Tübinge.
- Christ, O. (1994). A modular and flexible architecture for an integrated corpus query system. *arXiv preprint cmp-lg/9408005*.
- Christ, O. (1994). A modular and flexible architecture for an integrated corpus query system. *arXiv preprint cmp-lg/9408005*.
- Cook, V. (2005): *Second Language writing systems*. Clevedon: Multilingual Matters.
- Cook, V. J. (1997). L2 users and English spelling. *Journal of Multilingual and Multicultural Development*, 18(6), 474-488.
- Corder, S. P. (1993 [1973]): *Introducing Applied Linguistics*. Harmondsworth: Pelican.
- Dallapiazza, R.M./von Jan, E., Schönherr, T. (1998) (eds.): *Tangram: Deutsch als Fremdsprache. Kurs- und Arbeitsbuch 1 A*. Munich: Hueber.
- Cornish, F. (2009): Inter-sentential anaphora and coherence relations in discourse: a perfect match, in: *Language Science* 31 (2009), 572-592.
- Coulmas, F. (Hrsg.) (1986): *Direct and Indirect Speech*., Berlin, New York, Amsterdam.
- Crossley, S./Salsbury, T./McNamara, D. (2009): Measuring L2 Lexical Growth Using Hypernymic Relationships. In: *Language Learning* 59 (2) 307-334.
- Crossley, Sc.A./McNamara, D.S. (2011): Shared features of L2 writing: Intergroup homogeneity and text classification, in: *Journal of Second Language Writing* 20 (4) 271-285.
- Dale, E. (1965): Vocabulary measurement: Techniques and major findings. In: *Elementary English* 42, 895-901.
- Daller, H./Milton, J./Treffers-Daller, J. (eds.) (2007): *Modelling and Assessing Vocabulary Knowledge*. Cambridge: CUP.
- Daller, H./van Hou, R./Treffers-Daller, J. (2003): Lexical richness in spontaneous speech of bilinguals. In: *Applied Linguistics* 24, 197-222.
- Dellapiazza, R.M./von Jan, E., Schönherr, T. (1998) (Hrsgg.): *Tangram: Deutsch als Fremdsprache. Kurs- und Arbeitsbuch 1 A*. Munich: Hueber.
- Dewaele, J.-M. (2004): Individual differences in the use of colloquial vocabulary. The effects of sociobiographical and psychological factors. In: Bogaards, P./Laufer, L. (eds.): *Vocabulary in a second language*. Amsterdam: John Benjamins, 127-154.
- Dewaele, J.-M. (2004): The acquisition of sociolinguistic competence in French as a foreign language: An overview. In: *French Language Studies* 14, 301–319.
- Díaz-Negrillo, A./Fernández-Domínguez, J. (2006): Error-coding systems for learner corpora. In: *RESLA* 19, 83-102.

- Durrant, P./Schmitt, N. (2009): To what extent do native and non-native writers make use of collocations? In: *International Review of Applied Linguistics in Language Teaching (IRAL)*, 47 (2) 157-177.
- Durrant, P./Schmitt, N. (2009): To what extent do native and non-native writers make use of collocations? In: *International Review of Applied Linguistics in Language Teaching (IRAL)*, 47 (2) 157-177.
- Eckes, T. (2008): Rater types in writing performance assessments: A classification approach to rater variability. In: *Language Testing* 25 (2) 155-185.
- Eckes, T. (2009): *Reference Supplement to the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Section H: Many-Facet Rasch Measurement.* (http://www.coe.int/t/dg4/linguistic/manuel1_en.asp, December 2014.)
- Eisenberg, P. (2007): Sprachliches Wissen im *Wörterbuch der Zweifelsfälle*. über die Rekonstruktion einer Gebrauchsnorm. In: *Aptum. Zeitschrift für Sprachkritik und Sprachkultur* 3/2007: 209-228.
- Ellis, R. (1994): *The study of Second Language Acquisition*. Oxford: Oxford University Press.
- Fabricsius-Hansen, C. (2005): Elusive connectives. A case study on the explicitness dimension of discourse coherence. *Linguistics*, 43, 17-48.
- Fatemi, M. A. (2008) The relationship between writing competence, language proficiency and grammatical errors in the writing of Iranian tefl sophomores. Doctoral thesis. Universiti Sains Malaysia.
- Fender, M. Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language*, 2008, Volume 20, No. 1, pp. 19–42.
- Feng, L. (2010): Automatic readability assessment. PhD thesis, City University of New York (CUNY), <http://gradworks.umi.com/3426751.pdf> (February 2015).
- Foster, P./Tavakoli, P. (2009): Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. In: *Language Learning* 59 (4) 866-896.
- Fulcher, G. (2004): Deluded by Artifices? The Common European Framework and Harmonization. In: *Language Assessment Quarterly* 1 (4), 253-266.
- Fulcher, G./Davidson, F. (2007): *Language Testing and Assessment*. London/New York: Routledge.
- Gould, S.J. (1996): *The mismeasure of man*. London: Penguin.
- Gernsbacher, M. A./Givón, T. (Eds.) (1995): *Coherence in Spontaneous Text*. Benjamins, Amsterdam.
- Glaznieks, A. / Nicolas, L. / Stemle, E. / Abel, A. / Lyding, Verena (2014): Establishing a Standardised Procedure for Building Learner Corpora – a Response to Demands and Suggestions of Users. In: *Apples - Journal of Applied Language Studies* 8 (3), 2014 (<http://apples.jyu.fi/issue/view/15>)
- Graesser, A. C./Millis, K. K./Zwaan, R. (1997): Discourse comprehension, in: *Annual Review of Psychology* 48, 163-189.
- Granger, S. (2002): A Bird's-eye view of learner corpus research. In: Granger S./Hung, J./ Petch-Tyson, St (eds.): *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 3-33.
- Granger, S. (2003): Error-tagged learner corpora and CALL: a promising synergy. In: *CALICO Journal* 20 (3). Special issues on error analysis and error correction in computer-assisted language learning, 465-480.
- Granger, S. (2008): Learner corpora. In: Lüdeling, A. / Kytö, M. (eds.): *Corpus linguistics: an international handbook* (Handbooks of linguistics and communication science; 29.1_ 29.2). Berlin - New York: de Gruyter. 259-275.

- Granger, S., Bestgen, Y (2011) Categorizing spelling errors to assess L2 writing. *International Journal of Continuing Engineering Education and Life Long Learning*, 21, 2-3, 2011, 235 - 252.
- Guiraud, P. (1954): *Les caractères statistiques du vocabulaire*. Paris: Presse Universitaires de France.
- Gyllstad, H./Granfeldt, J./Bernardini, P./Källkvist, P. (2014): Linguistic correlates to communicative proficiency levels of the CEFR. The case of syntactic complexity on written L2 English, L3 French and L4 Italian. In: *EUROSLA Yearbook 14*, 1-30.
- Halliday, M. A. K. /Hasan, R. (1989): ***Language, context and text: a social semiotic perspective***. Oxford: Oxford University. Press.
- Halliday, M. A. K./Hasan, R. (1976): *Cohesion in English*. London, Longman.
- Hana, J./Rosen, A./ Štindlová, B./Štěpánek, J. (2014): Building a learner corpus. In: *Language Resources and Evaluation 8*, SE
- Hana, J./Rosen, A./Skodová, S. Stindlová, B.: Error-tagged learner corpus of Czech. In: *Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010, Uppsala, Sweden 2010*, 11-19.
- Hancke J./Meurers D./Vajjala S. (2012): Readability Classification for German using lexical, syntactic, and morphological features. In: ***Proceedings of the 24th International Conference on Computational Linguistics (COLING)***, 1063-1080.
- Hancke, J. ***Automatic Prediction of CEFR Proficiency Levels Based on Linguistic Features of Learner Language***. Master's thesis, Universität Tübingen, April 2013.
- Hancke, J./Meurers, D./Vajjala, D. (2012): Readability classification for German using lexical, syntactic, and morphological features. In: *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*, 1063-1080, Mumbai, India.
- Hasil, J./Hájková, E./Hasilová, H. (2007): ***Brána jazyka českého otevřená***. Prague: Karolinum.
- Hasko, V. (2013): Capturing the Dynamics of Second Language Development via Learner corpus research: a very long engagement. In: *The Modern Language Journal 97*, S1, 1-10.
- Hattingh, K. (2005) The syntactic development of grade 12 ESL learners. Dissertation, North-West University, Potchefstroom Campus.
- Hawkey, R./Barker, F. (2004): Developing a Common Scale for the Assessment of Writing. In: *Assessing Writing 9*, 122-159.
- Hawkins, J. A./Filipović, L. (2012): ***Criterial features in L2 English: Specifying the reference levels of the Common European Framework***. Cambridge: CUP.
- Held, G. (1995): *Verbale Höflichkeit. Studien zur linguistischen Theorienbildung und empirische Untersuchung zum Sprachverhalten französischer und italienischer Jugendlicher in Bitt- und Dankessituationen*. Tübingen: Narr.
- Holmes/Brown 1976: Developing sociolinguistic competence in a second language. *Tesol Quarterly 10/4*.
- Housen, A., Kuiken, F. (2009) Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics*, 30(4): 461-473
- Housen, A./Kuiken, F. (2009): Complexity, Accuracy, and Fluency in Second Language Acquisition. In: *Applied Linguistics 30 (4)* 461-473.
- Housen, A./Kuiken, F./Vedder, I. (eds.) (2012): ***Dimensions of L2 Performance and Proficiency Investigating Complexity, Accuracy and Fluency in SLA***. Amsterdam: John Benjamins.

- Hovermale, DJ., Martin, S. Developing an Annotation Scheme for ELL Spelling Errors. Proceedings of MCLC-5 (Midwest Computational Linguistics Colloquium) East Lansing, Michigan, USA, May 10-11, 2008.
- Hudson, Tom/Detmer, Emily/Brown, J.D. (1992): A framework for testing cross-cultural pragmatics. Hawaii: University of Hawai'i at Manoa.
- Hudson, Tom/Detmer, Emily/Brown, J.D. (1995): *Developing prototypic measures of cross-cultural pragmatics*. Hawaii: University of Hawai'i at Manoa.
- Hulstijn, J. H. (2007): The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. In: *The Modern Language Journal* 91, 663-667.
- Hulstijn, J. H./Alderson, C./Schoonen, R. (2010): Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them? In: Bartning, I./Martin, M./Vedder, I. (eds.): *Communicative Proficiency and Linguistic Development: intersections between SLA and language testing research*. Eurosla Monograph Series. (<http://eurosla.org/monographs/EM01/EM01home.html>)
- Hymes, D. (1974): *Foundations in Sociolinguistics*. Philadelphia: University of Pennsylvania Press.
- Ishikawa, S. (1995) Objective Measurement of Low-Proficiency EFL Narrative Writing. *Journal of Second Language Writing*, 4: 51 - 70.
- Jarvis, S. (2002): Short texts, best-fitting curves and new measures of lexical diversity. In: *Language Testing* 19 (1) 57-84.
- Jelínek, T., Barbora Štindlová, Alexandr Rosen, Jirka Hana (2012). Combining Manual and Automatic Annotation of a Learner Corpus. Text, Speech and Dialogue Lecture Notes in Computer Science Volume 7499, pp 127-134 <http://ufal.mff.cuni.cz/~hana/bib.html#rosen-et-al-2013-czesl-lre>, December 2014.
- Johns, T. (1988): Whence and whither classroom concordancing? In: Bongaarts, T./de Haan, P./Lobbe, S./Wekker, H. (eds.): *Computer Applications in Language Learning*. Dordrecht: Foris, 9-33.
- Johns, T. (1997): Contexts: The Background, Development and Trialling of a Concordance-based CALL Program. In: Wichmann, Anne/Fligelstone, Steven/McEnery, Tony/Knowles, Gerry (eds.) (1997), *Teaching and Language Corpora*. London: Longman, 100-115.
- Kaczmarek, C. M. (1980) Scoring and Rating 'essay tasks'. (& Oller, J.W. and Perkins, K. &. *Research in Language Testing*. Rowley, Massachusetts: Newbury House.)
- Koch, P./Oesterreicher, W. (2011²) : *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Berlin u.a.: de Gruyter.
- Kroll, B. (1990) What does time buy? ESL student performance on home versus class compositions. In B. Kroll, ed. *Second language writing: Research insight for the classroom*. Cambridge: CUP, pp. 140 – 154.
- Kuiken, F./Vedder, I./Gilbert, R. (2010): Communicative Adequacy and Linguistic complexity in L2 writing. *EUROSLA Monographs Series* 1, 81-100.
- Larsen-Freeman, D. (2009): Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. In: *Applied Linguistics* 30 (4) 579-589.
- Laufer, B. (1995): Beyond 2,000. A measure of productive lexicon in a second language. In: Eubank, L./Selinker, L./Sharwood Smith, M. (eds.): *The Current State of Interlanguage*. Amsterdam/Philadelphia: John Benjamins, 265-272.
- Laufer, B./Nation, P. (1995): Vocabulary size and use: lexical richness in L3 written production. In: *Applied Linguistics* 16, 307-322.
- Linnarud, M. (1986): *Lexis in composition: A performance analysis of Swedish learners' written English*. Malmö: CWK Gleerup.

- Little, D. (2007): The Common European Framework of Reference for Languages: Perspectives on the Making of Supranational Languages Education Policy. In: *The Modern Language Journal* 91, 645-655.
- Lorenzo-Dus, N. (2007): The best of both worlds? Combined methodological approaches to the assessment of vocabulary in oral proficiency interviews. In: Daller/Milton/Treffers-Daller (eds.) 220-233.
- Louwerse, M.M./Graesser, A. C. (2004). Coherence in discourse, in: Strazny, P. (Ed.): *Encyclopedia of linguistics*. Chicago: Fitzroy Dearborn.
- Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics* 14, 3–28(26). URL <http://www.ingentaconnect.com/content/jbp/ijcl/2009/00000014/00000001/art00002>.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4):474-496.
- Lu, X. (2011): A corpus-based evaluation of syntactic complexity measures as indices of College-level ESL writers' language development. In: *TESOL Quarterly* 45 (1) 36-62.
- Lu, X. (2012): The relationship of lexical richness to the quality of ESL learners' oral narratives. In: *The Modern Language Journal*, 190-208.
- Lüdeling, A. (2008): Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora. In: Walter, M./Grommes, P. (eds.): *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitsprachenerwerbsforschung*. Tübingen: Niemeyer, 119-140.
- Lüdeling, A./Walter, M./Kroymann, E./Adolphs, P. (2005): Multi-level Error Annotation in Learner Corpora. In: Hunston, S./Danielsson, P. (eds.): *Proceedings from the Corpus Linguistics Conference Series* (Corpus Linguistics 2005, Birmingham, 1415 July 2005). (<http://www.corpus.bham.ac.uk/PCLC>).
- Lyster, R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' sociolinguistic competence. *Applied Linguistics*, 15 (3) 263–287.
- Malvern, D./Richards, B./Chipere, N./Durán, P. (2008²): *Lexical Diversity and Language Development. Quantification and Assessment*. New York: Palgrave Macmillan.
- Matsuda, P. K. (1997): Contrastive rhetoric in context: A dynamic model of L2 writing, in: *Journal of Second Language Writing*, 6, 45-60.
- McCarthy, P.M./Jarvis, S. (2007): vocd: A theoretical and empirical evaluation. In: *Language Testing* 24 (4) 459-488.
- McNamara, D. S./Kintsch, E./Butler Songer, N./Kintsch, W. (1996): Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning from Text, in: *Cognition and Instruction*, Vol. 14, No. 1 (1996), 1-43.
- McNamara, D.S./Kintsch, W. (1996): Learning from text: Effects of prior knowledge and text coherence, in: *Discourse Processes*, 22, 247-287.
- McNamara, D.S./Louwerse, M.M. /Graesser, A.C. (unpublished): Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension. Grant proposal.
- Mellor, A. (2011): Essay Length, Lexical Diversity and Automatic Essay Scoring. In: *Memoirs of the Osaka Institute of Technology*, Series B Vol. 55, No. 2 (2011), 1-14.
- Ménard, N. (1983): *Mesure de la richesse lexicale. Théorie et vérifications expérimentales. Études stylométriques et sociolinguistiques*. Genf/Paris: Slatkine-Champion.
- Meurers, D. (2012): Natural Language Processing and Language Learning. *Encyclopedia of Applied Linguistics*. Blackwell. purl.org/dm/papers/meurers-11.html.

Mezzadri, M. (2000): *Rete! Book 1*. Perugia: Guerra Edizioni.

Mougeon, R./Dewaele, J.-M. (2004): Preface. In: IRAL 42 (4) (Special Issue: Variation in the interlanguage of advanced second language learners.), 295-301.

5.8.1 Mougeon, Raymond/Nadasdi, Terry/Rehner, Katherine (2010): *The Sociolinguistic Competence of Immersion Students*.

Müller, Ch./Strube M. (2006): Multi-Level Annotation of Linguistic Data with MMAX2. In: S. Braun, K. Kohn, J. Mukherjee (Eds.): *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods*. Frankfurt: Peter Lang, 197-214.

Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal*, 87, 261-276.

Nation, P. (2001): *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, P. (2007): Fundamental issues in modelling and assessing vocabulary knowledge. In: Daller, H./ Milton, J./Treffers-Daller, J. (eds.): *Modelling and Assessing Vocabulary Knowledge*. Cambridge: Cambridge University Press.

Nesselhauf, N. (2005): *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
North, B. (2000): *The Development of a Common Framework Scale of Language Proficiency*. Oxford: Peter Lang.

North, B. (2000): *The Development of a Common Framework Scale of Language Proficiency*. Oxford: Peter Lang.

Nuzzo, E. (2007): Imparare a fare cose con le parole. Richieste, proteste, scuse in italiano lingua seconda. Perugia: Guerra.

Nuzzo, E. (2009): "Buongiorno, ho bisogno dell'informazione per andare a Barcellona": uno studio longitudinale sulle richieste di informazioni e suggerimenti in italiano L2. In: *Linguistica e Filologia* 28, 83 –109.

O'Loughlin, K. (1995): Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test. In: *Language Testing* 12 (2) 217-237.

Okada, T. A Corpus Analysis of Spelling Errors Made by Japanese EFL Writers.

Ortega, L. (2003) Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24 (4), 492-518.

Ortega, L. (2012): Interlanguage complexity: A construct in search of theoretical renewal. In: Szmrecsanyi, B./Kortmann, B. (Eds): *Linguistic complexity in interlanguage varieties, L2 varieties, and contact languages*. Berlin: Walter de Gruyter.

Ott, N. (2009). Information Retrieval for Language Learning: An Exploration of Text Difficulty Measures. Master's thesis, University of Tübingen, Seminar für Sprachwissenschaft, Tübingen, Germany. URL <http://drni.de/zap/ma-thesis>.

Paquot, M./Granger, S. (2012): Formulaic language in Learner Corpora. In: *Annual Review of Applied Linguistics* 32, 130-149.

Perfetti, C. A., Rieben, L. & Fayol, M. (eds) (1997). *Learning to Spell: Research, Theory, and Practice across Languages*. Mahwah, NJ: Lawrence-Erlbaum Associates.

Petersen, S.E./Ostendorf, M. (2009): A machine learning approach to reading level assessment. In: *Computer Speech and Language* (23) 86-106.

Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.

- Pollitt, A./Murray, N.L. (1996): What raters really pay attention to. In: Milanovic, M./Saville, N. (eds.): *Performance testing, cognition and assessment; Selected papers from the 15th Language Testing Research Colloquium*. Cambridge: Cambridge University Press, 74-91.
- Preston, D. (2000). Three kinds of sociolinguistics and SLA: A psycholinguistic perspective. In B. Swierzbis, F. Morris, M. E. Anderson, C. E. Klee and E. Tarone (eds), *Social and cognitive factors in second language acquisition*. Somerville: Cascadia Press, pp. 3–30.
- Read, J. (2000): *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Read, J. (2007): Second language vocabulary assessment: current practice and new directions. In: *International Journal of English Studies* 7 (2) 105-125.
- Read, J./Chapelle, C. (2001): A framework for second language vocabulary assessment. In: *Language Testing* 18, 1-32.
- Read, J./Nation, P. (2004): Measurement of formulaic sequences. In: Schmitt, N. (ed.): *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins, 23-35.
- Regan, V. (1995). The acquisition of sociolinguistic native speech norms. In: B. Freed (ed.), *Second language acquisition in a study abroad context*. Amsterdam, Philadelphia: Benjamins, pp. 245–267. (vorgemerkt)
- 5.9 Regan, Vera/Howard, Martin/Leme, Isabelle (2009): *The Acquisition of Sociolinguistic Competence in a Study Abroad Context*.
- Reznicek, M./Lüdeling, A./Hirschmann, H. (in print): Competing Target Hypotheses in the Falko Corpus. A Flexible Multi-Layer Corpus Architecture. In: Díaz-Negrillo, A./Ballier, N./Thompson, P. (eds.): *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins (Series Studies in Corpus Linguistics).
- Reznicek, M./Lüdeling, A./Krummes, C./Schwanteschke, F./Walter, M./Schmidt, K./Hirschmann, H./Andreas, T. (2012): *Das Falko-Handbuch. Korpusaufbau und Annotationen*. Version 2.01. HU Berlin (http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko/Falko-Handbuch_Korpusaufbau%20und%20Annotationen_v2.01)
- Rimrott, A., Heift, T. (2008). Evaluating automatic detection of misspellings in German. *Language Learning & Technology*.
- Römer, U. (2010): Using general and specialized corpora in English language teaching: past, present and future. In: Campoy-Cubillo, M. et al. (eds.): *Corpus-based approaches to English Language Teaching*. London: Continuum, 18-38.
- Römer, Ute. 2008. 7. Corpora and language teaching. In: Lüdeling, Anke & Merja Kytö (eds.). *Corpus Linguistics. An International Handbook (volume 1)*. [HSK series] Berlin: Mouton de Gruyter. 112-130.
- Römer, U. (2006): Pedagogical applications of corpora: some reflections on the current scope and a wish list for future developments. In: *Zeitschrift für Anglistik und Amerikanistik* 54 (2) 121-134.
- Rose, K.R. (2000), An exploratory cross-sectional study of interlanguage pragmatic development. *Studies in Second Language Acquisition* 22, 1, 27-67.
- Rosen, A., Jirka Hana, Barbora Štindlová, and Anna Feldman (2013): Evaluating and automating the annotation of a learner corpus. *Language Resources and Evaluation*, pages 1-28, April 2013. <http://ufal.mff.cuni.cz/~hana/bib.html#rosen-et-al-2013-czesl-lre>, December 2014.
- Sanders, T./Pander Maat, H. (2006): Cohesion and coherence: Linguistic approaches, in: Brown, K., et al. (Eds.), *Encyclopedia of Language and Linguistics*. Elsevier, London.
- Sassoon, R. (1995). *The Acquisition of a Second Writing System*. Oxford: Intellect.

- Schmitt, N. (ed.) (2004): *Formulaic Sequences: Acquisition, Processing, and Use*. Amsterdam: John Benjamins Press.
- Schmitt, N./Carter, N. (2004): Formulaic sequences in action: An Introduction. In: Schmitt, N. (ed.): *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins, 1-21.
- Schneider, G./North, B. (2000): *Fremdsprachen können - was heißt das? Skalen zur Beschreibung, Beurteilung und Selbsteinschätzung der fremdsprachlichen Kommunikationsfähigkeit*. Nationales Forschungsprogramm 33, Wirksamkeit unserer Bildungssysteme. Chur, Zürich: Rüegger.
- Schneider, J. G. (2013): Sprachliche ‚Fehler‘ aus sprachwissenschaftlicher Sicht. In: *Sprachreport* 1-2/2013, 30-37.
- Skehan, P. (2009): Modelling Second Language Performance: Integrating complexity, accuracy, fluency and lexis. In: *Applied Linguistics* 30 (4) 510-532.
- Škodová, S., Barbora Štindlová, Jirka Hana and Alexandr Rosen (2011). Víceúrovňová anotace českého žákovského korpusu. In: V. Petkevič and A. Rosen (Eds.), *Korpusová lingvistika Praha 2011: 3 - Gramatika a značkování korpusů*, 16. pp 208-225. Studie z korpusové lingvistiky. Nakladatelství Lidové noviny: Praha.
- Spinelli, B./Parizzi, F. (ed.) (2010): *Profilo della lingua italiana*. Firenze: La Nuova Italia.
- Stede, M. (2007): Korpusgestützte Textanalyse. Grundzüge der Ebenen-orientierten Textlinguistik. Tübingen: Narr.
- Spooren, W./Sanders, T. (2008): The acquisition order of coherence relations: On cognitive complexity in discourse, in: *Journal of Pragmatics* 40 (2008), 2003-2026.
- Štindlová B., S. Škodová, A. Rosen and J. Hana (2012). Annotating foreign learners' Czech In: *Studies in Formal Slavic Linguistics. Contributions from Formal Description of Slavic Languages 8.5*, Peter Lang GmbH, Frankfurt am Main, Germany, pp. 205-219.
- Taguchi, N. (2012): *Context, individual differences and pragmatic competence*. Bristol: Multilingual Matters.
- Tarone, E. (2007): Sociolinguistic approaches to second language acquisition research, 1997-2007. In: *Modern Language Journal* 91, 837-848.
- Trosborg, A. (1995), *Interlanguage pragmatics: Requests, complaints, apologies*. Berlin: Mouton de Gruyter.
- UIMA Framework: <http://uima.apache.org> (July 2013)
- Ure, J.N. (1971): Lexical density and register differentiation. In: Perren, G.E./Trim, J.L.M. (eds.): *Applications of Linguistics: Selected papers of the Second International Congress of Applied Linguistics*. Cambridge: CUP, 443-452.
- Vajjala, S./Meurers, D. (2012): On improving the accuracy of readability classification using insights from second language acquisition. In: Joel Tetreault, Jill Burstein, and Claudial Leacock, editors, *Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA7) at NAACL-HLT*. Montreal, Canada, June 2012, 163-173. <http://aclweb.org/anthology/W12-2019.pdf> (February 2015).
- van Compernelle, Remi A. /Williams, Lawrence (2012): Teaching, Learning, and Developing L2 French Sociolinguistic Competence: A Sociocultural Perspective *Applied Linguistics* (2012) 33(2): 184-205 first published online January 24, 2012.
- Vaughan, C. (1991): Holistic assessment: What goes on in the rater's mind? In: Hamp-Lyons L. (ed.): *Assessing Second Language Writing in Academic Contexts*. Norwood: Ablex, 111.125.
- Vedder, I.(2007): Competenza pragmatic e complessità sintattica in italiano L2: l'uso dei modificatori nelle richieste. In: *Linguistica e Filologia* 25, 99-123.

- Vivanco, V. (2005): The absence of connectives and the maintenance of coherence in publicity texts, in: *Journal of Pragmatics* 37 (2005), 1233-1249.
- Warga, M. and U. Scholmberger (2007). The acquisition of French apologetic behaviour in a study abroad context. *Intercultural Pragmatics*, 4, 221-251.
- Wesche, M./Paribakht, T.S. (1996): Assessing second language vocabulary knowledge depth versus breadth. In: *The Canadian Modern Language Review* 53, 13-40.
- Wisniewski, K. (2010): Bewertervariabilität im Umgang mit GeRS-Skalen. Ein- und Aussichten aus einem Sprachtestprojekt. In: *Deutsch als Fremdsprache* 3, 143-150.
- Wisniewski, K. (2012): Lexikalische Kompetenzen in der Fremdsprache testen: Ein Modellierungsansatz. In: In: Abel, A. / Vettori, C. / Wisniewski, K. (eds.): *Gli studenti altoatesini e la seconda lingua: indagine linguistica e psicosociale. / Die Südtiroler SchülerInnen und die Zweitsprache: eine linguistische und sozialpsychologische Untersuchung. Volume 2 – Band 2. Bolzano - Bozen: Eurac. , 24-49. (<http://www.eurac.edu/en/research/publications/PublicationDetails.aspx?pubId=0100156&type=Q>)*
- Wisniewski, K. (2013): The empirical validity of the CEFR fluency scale: the A2 level description. In: Galaczi, E.D./Weir, C.J. (eds.): *Exploring Language Frameworks: Proceedings of the ALTE Krakow Conference*. Cambridge: Cambridge University Press, 253-272. *Studies in Language Testing*.
- Wisniewski, K. (2014): ***Die Validität der Skalen des Gemeinsamen europäischen Referenzrahmens für Sprachen. Eine empirische Untersuchung der Flüssigkeits- und Wortschatzskalen des GeRS am Beispiel des Italienischen und des Deutschen***. Frankfurt: Peter Lang. *Language Testing and Evaluation Series*, 33.
- Wisniewski, K. / Abel, A. (2012): Die Sprachkompetenzerhebung: Theorie, Methoden, Qualitätssicherung. In: Abel, A. / Vettori, C. / Wisniewski, K. (eds.): *Gli studenti altoatesini e la seconda lingua: indagine linguistica e psicosociale. / Die Südtiroler SchülerInnen und die Zweitsprache: eine linguistische und sozialpsychologische Untersuchung. Volume 1 – Band 1. Bolzano - Bozen: Eurac. 13-64 (<http://www.eurac.edu/en/research/publications/PublicationDetails.aspx?pubId=0100156&type=Q>)*
- Wisniewski, K./Schöne, K./Nicolas, L./Vettori, C./ Boyd, A./Meurers, D./ Abel, A./Hana, J. (2013): MERLIN: An online trilingual learner corpus empirically grounding the European Reference Levels in authentic learner data. In: ***ICT for Language Learning, Conference Proceedings 2013***. Libreriauniversitaria.it Edizioni. (http://conference.pixel-online.net/ICT4LL2013/common/download/Paper_pdf/322-CEF03-FP-Wisniewski-ICT2013.pdf)
- Wolfe-Quinteiro, K., Inagaki, S., Kim, H.-Y. (1998) Second language development in writing: measures of fluency, accuracy, & complexity. Hawai'i: University of Hawai'i at Manoa.
- Wray, A. (2002): *Formulaic Language and the Lexicon*. Cambridge: CUP.
- Yang, W./Sun, Y. (2012): The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. In: ***Linguistics and Education***, 23 (1), 31-48.
- Wray, A. (2002): ***Formulaic Language and the Lexicon***. Cambridge: Cambridge University Press.
- Young, R. (1999). Sociolinguistic approaches to SLA. *Annual Review of Applied Linguistics*, 19: 105–32.
- Yu, Ming Chung (2005): Sociolinguistic Competence in the Complimenting Act of Native Chinese and American English Speakers: A Mirror of Cultural Value. In: *Language and Speech* 48 (1) 91-119.
- Yu, Ming-chung (2011): Learning how to read situations and know what is the right thing to say or do in an L2: A study of socio-cultural competence and language transfer. *Journal of Pragmatics*, 2011, Vol.43(4), psp.1127-1147.
- Zeldes, A./Ritz J./Lüdeling A. et al. (2009): ***Annis: A search tool for multi-layer annotated corpora. In Proceedings of Corpus Linguistics***, July 20-23. Liverpool. (<http://ucrel.lancs.ac.uk/publications/cl2009/>).

Zipser, F./Romary, L./al. (2010). A model oriented approach to the mapping of annotation formats using standards. In: *Workshop on Language Resource and Language Technology Standards, LREC 2010*.

Zuskin, Robin D. (1992): Assessing L2 Sociolinguistic Competence: In Search of Support from Pragmatic Theories. In: *Pragmatics and Language Learning 4*.