



MERLIN ANNOTATION SCHEME[©]

The Annotation scheme can be freely downloaded.

Please cite as: Katrin Wisniewski, Claudia Woldt, Karin Schöne, Andrea Abel, Verena Blaschitz, Barbara Štindlová, Kateřina Vodičková (2014):
The MERLIN annotation scheme for the annotation of German, Italian, and Czech learner language . Available online www.merlin-platform.eu



This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

MERLIN ANNOTATION SCHEME (I)

GRAMMAR TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
word order	word order in main clause	G_Wo_womaincl	This tag is used if an error occurs in a main clause in the following cases: a) verb in the wrong position b) wrong order of syntactical items such as attribute, subject, object. In some cases, both the position of the verb AND the position of subject, object etc. are wrong. No differentiation is made in this case. The tag also applies to word order errors in questions and exclamations. LANGUAGE-SPECIFIC REMARKS GER: includes errors in the use of "zu" (only wrong position, not missing!) CZE: includes errors in the use of "se"/"si" (only wrong position, not missing!) RELATED TAGS If a participle, auxiliary, or copula is in the wrong position, it is tagged as G_Verb_compl_Pos.	wrong clause including punctuation mark	a) *[Včera učil jsem se česky.], *[Vielleicht du könntest mir bei meine Wohnungssuche helfen.], *[Sollst du Wasser und Bikini mitbringen.]; *[Per trovarlo ti vorrei aiutare.] b) *[Známe film český.], *[Učil se jsem česky.], *[Könntest du bitte mich anrufen?], *[Könntest Du bitte mir helfen bei der Wohnungssuche?]; *[Mia famiglia sta anche bene.], *[Vorrei andare a ti trovare.]	R/T/U
word order	word order in subordinate clause	G_Wo_wosubcl	This tag is used if an error occurs in a subordinate clause in the following cases: a) verb in the wrong position b) wrong order of syntactical items such as attribute, subject, object. In some cases, both the position of the verb AND the position of subject, object etc. are wrong. No differentiation is made in this case. The tag also applies to word order errors in questions and exclamations.	wrong clause including punctuation mark	a) První je těžký znát, [co učil se.], *Ich wollte wissen, [welche Sprache wird in der Familie gesprochen.], *wenn haben Sie Zeit.], dann bitte sagen Sie mir. *Gerne würde ich auch wissen [welche Aufenthaltsformalitäten verlangen Sie von mir?] b) Maminka myslí, [že musí pracovat moc pěkně pro lepší budoucnost.], *Ich habe nicht verstanden, [von welche fährt der Zug Bahnhof.]	R/T/U
negation	general negation error	G_Neg_neggen_Pos G_Neg_neggen_O G_Neg_neggen_Ch G_Neg_neggen_Ad	This tag is used in case of a) wrong position of negation expressions (G_Neg_neggen_Pos) b) missing part of negation (G_Neg_neggen_O) c) wrong use of negation words (e.g., GER: nicht, nein, kein; ITA: no, non; CZE: ne, žádný (G_Neg_neggen_Ch) d) redundant/wrongly added negation word (G_Neg_neggen_Ad)	a) _pos: 1 token; 2 tokens for CZE (neg. word and verb, if neg. word wrongly distributed) b) _o: 1 or more token - the POS to be negated; c) _ch: 1 token - the negation word	a) *[mám ne] kávu, *[půjdu neráno], *Ich habe Hunger [keinen]; *Io credo [non]. b) *On [ne] velký; • [Io mangio nè carne nè pesce] {Io non mangio né carne né pesce}; *[Luca non va a scuola perché ne ha voglia; {Luca non va a scuola perché non ne ha voglia} *{Non è né chiaro scuro} {Non è né chiaro né scuro}. c) *Bohužel, nemám [ne] čas; *Ich habe [nicht] Zeit. *Er wird dort arbeiten [nein]; *[Non], viene più tardi. d) *Man kann nicht auf solchen grössen Teil seiner Persönlichkeit zu verzichten, ohne seine psychische Gesundheit [nicht] zu schaden.	R
negation	double negation	G_Neg_negdoub	This is a single-language tag that is only applied for CZECH. In Czech, all negated pronouns require a negated verb form. This tag refers to the missing negation particle "ne" at the verb and vice versa.	1 token (verb, existing part of the double negation)	*[mám] žádný čas {nemám žádný čas} *nikdo [volal] → {nikdo nevolal} *nemám [nějaký] čas {nemám žádný čas }	R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
verb valency (obligatory arguments)	complement number	G_Valency_complnumb_O G_Valency_complnumb_Ad	In Merlin, the erroneous <i>presence or absence</i> of obligatory arguments is annotated. Free complements (Helbig: "freie Angaben") are not tagged. The tags are specified as follows: a) omission of a subject or obligatory object (G_Valency_complnumb_O) b) wrong addition of a subject or obligatory object (G_Valency_complnumb_Ad). This tag was used as scarcely as possible. Annotators were asked to rely on their native speaker competence to decide whether an explicit subject really is an error (annotated) or a question of style (not annotated). Of course, obligatory arguments may be missing because the learner gives the impression to have created an elliptic construction consciously. In those cases, the missing argument is not tagged. A redundant subject may occur in Italian and Czech. Often, the use of an explicit subject in these languages is subjective and does not necessarily represent an error. The tag was used only in the case of an obvious error clearly contradicting target language usage norms. LANGUAGE-SPECIFIC REMARKS GER: "Es" as correlate to an obligatory or facultative complement may be wrongly added or absent. For both, the G_Valency_complnumb tag is used. RELATED TAGS The omission of a subject or obligatory object usually leads to a reference problem. The coherence/cohesion tag (C_coh_ref) was then used additionally. If a learner omits a clitic that is an obligatory object, G_Clit_O is used, too, e.g. *Ho dimenticato di scrivere [scriverlo] prima.	wrong argument (usually 1 token) or whole clause including punctuation mark (if argument is missing)	a) *Já vstávám v 5, [ale vstáváš v 8.] {...ale ty vstáváš v 8.} *Ich liebe {dich}. *Er hat uns nicht gesagt, ob {er} kommen will. *[Spero che possa aiutarLa.]{Spero che questo possa aiutarLa.}, *[Forse con cui si parla in chat non è quella persona che tu pensi] {Forse la persona con cui si parla in chat non è quella persona che tu pensi}." b) *Petr vstává v 6 hodin. On nesnídá, protože [on] nemá hlad.; *Susanne denkt [sie] zu viel nach. *Che lavoro [tu] fai adesso? *Secondo me [tu] hai lavorato. *Che cosa di preciso [tu] vuoi fare?, * Adesso [la mia professione] sono insegnante.	R
agreement (subject and verb)		G_Agr	An agreement error in MERLIN includes wrong grammatical forms in the following combinations of subject and verb (incl. copulative predicates): a) wrong form (person, number) of the verb b) wrong form (person, number) of an auxiliary or participle c) wrong part of speech (POS) in copulative predicate (infinitive instead of adjective etc.) d) wrong form (person, number) of copula Of course, combinations of adjective and substantive, pronoun and substantive, numerals and substantives etc. may contain agreement errors, too. In Merlin, though, it was decided to annotate these as morphological errors (see G_Morphol tags). This way, the difficult categorization of a phenomenon as either an agreement error or a complement error was avoided. RELATED TAGS If the subject is not explicit, but inferable from the context, and the verb form is grammatically correct, no G_Agr tag is applied. The phenomenon is tagged as a coherence problem (TH2 & EA2, C_coh_ref). Examples: *Secondo che tipo lavoro vuole fare? *Come stai? Che cosa fai? Ha anche molto lavoro? (In both cases the subject suggested by the context is "tu" and the speaker was addressing it informally in the rest of the text).	wrong form (usually 1 token)	I) a) *Jana [čtu], *Petr [nerozumět]; *chci se [učí]; *Jana [lese], *Peter [verstehen]; *Gianni [leggo], *Chiara [capire]; *Noi [volevano] un gelato; *Tu [sapeva] questo. b) *Jana [jsi] udělala, *přišel [jsme]; *[abychom] mohl {abych mohl}; *Jana [has] gelesen, *wie [seid] gekommen; *Gianni [abbiamo] letto, *(noi) [sono] venuti, *Siamo potuti (abbiamo potuto) giocare c) *Petr je [unavit]; *Jana ist [müde werden]; *Chiara è [stancare] d) *Petr [jsou] unavený; *Jana [sind] müde; *Giovanni e Luca [siamo] stanchi	R/C
reflexivity	reflexive pronoun	G_Refl_pronrefl_O G_Refl_pronrefl_Ad G_Refl_pronrefl_Ch G_Refl_pronrefl_Pos	Errors in the use of reflexive pronouns in MERLIN are tagged in case of: a) missing pronoun (G_Refl_pronrefl_O) b) wrongly added reflexive pronoun (G_Refl_pronrefl_Ad) c) wrong choice of reflexive pronoun (also in the object position) or using personal pronoun instead of the reflexive one (G_Refl_pronrefl_Ch) d) wrong position of the reflexive pronoun (G_Refl_pronrefl_Pos) RELATED TAGS In case of errors in reflexives in the object position, it is hard to decide whether it is an agreement error (G_Agr) or an error in choice of the right pronoun. In this cases we tag the error as reflexivity error.	wrongly positioned reflexive pronoun or token containing the reflexive pronoun, in case of missing reflexive: verb (verb form)	a) *[dívám] na televizi, *ráno [maluju] {instead of: ráno se maluju}, *er [entschuldigt], *Laura und Ferdinand reden [sich]; *Michele [rivolge] à sua sorella b) *dnes [si] neučí, *vidím [se] Petra; *[mi] desidero una camera singola c) *myju [se] ruce, *myslí [se] to, *smála [si]; *abychom [nás] mohly vidět abychom {se} mohly vidět; *[se] {si} lava ogni mattina, *[si] {ci} rivolgiamo a Lei d) Já těším na Tvou oslavu. Chtěla bych zeptat se., *[mi] non posso incontrare con tu	R/T
reflexivity	possessive reflexive pronoun	G_Refl_pronreflposs	This is a single-language tag that is only applied for CZECH. Czech distinguishes between non-reflexive possessive pronouns (můj, tvůj...) and reflexive possessive pronouns (svůj, svoje...). Confusion between these are annotated with G_Refl_pronreflposs RELATED TAGS If there is also a form problem (*stojí před její dům instead of *svým domem), the pronoun will be tagged as agreement error (G_Agr) and as G_Refl_pronreflposs	wrong reflexive pronoun	*potřebuju [moji] knihu, * vidím [mého] otce	R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
inexistent inflection	adjective inflection	G_Infect_adj_inexist	This tag is used if an adjective form contains errors, i.e. if the learner uses an inflected form of the adjective that does not exist in the inflectional paradigm at all. The tag includes wrong forms of comparative/ superlative, too. LANGUAGE-SPECIFIC REMARK GER: Beyond that, the tag is used if the learner uses the determinate paradigm instead of the indeterminate. In this case, the form may exist but it belongs to another paradigm. RELATED TAGS If the form exists but belongs to another case or number, the G_Morphol tags are used.	wrong adjective form (1 or more token)	GER: *Das ist ein schöne Haus. (determinate instead of indeterminate) *ein [blaus] Himmel {blauer}; *... mit [freundlich] Grüßen ITA: *una persona [pedanta] {pedante}, *un tempo [mutevolo] {mutevole}; *polli [arrosti] {polli arrosto}; *poltrone [lille] {poltrone lilla} CZE: v [následujícím] {následujícím} měsíci, s [cizou] {cizi} nabídkou	R
inexistent inflection	noun inflection	G_Infect_noun_inexist	Generally, substantives, articles, adjectives, numerals, and pronouns count as nouns. In Merlin, we decided to separate errors in the inflection of adjectives from all other types of nouns. Therefore, this tag is used if the form of a substantive, article, numeral or pronoun contains an error. It only applies if the learner uses an inflected form that does not exist in the inflectional paradigm at all. RELATED TAGS If the form exists but simply belongs to another case or number, the G_Morphol tags are used. If a learner writes, e.g. *un'aiuta instead of "un aiuto" or *passeggiati instead of "passeggiate", we tag G_Infect_noun and not V_word/fs_nonexist . The reason is that we assume that the main problem is not that the word does not exist but that the learner has chosen a wrong gender which is a form that does not exist in the paradigm of the noun. The same applies for with G_Infect_verb and G_Infect_adj . Furthermore, in cases like *un'aiuta vs "un aiuto", neither the article (G_Art) nor the apostrophe error (O_Apostr) are tagged. These errors depend on the gender error of the noun, while the learner was consistent in the (erroneous) form used.	wrong noun form (1 token)	GER: *[euche] [Fahrrade], *er gab es [mire] {mir}; * welche Rechte und [Pfichte]... ITA: *Con [personi] {persone}, *chiedere [un'aiuta] {un aiuto}, *[un Suo Risposto] {una Sua risposta} CZE: dej to [mojemu] {mému} kamarádovi, promluvit k [lidům] {lidem}, o [ticicích] {tisících} nápadů	R
inexistent inflection	verb inflection	G_Infect_verb_inexist	If the learner uses a verb form that does not exist in the paradigm of the verb, this tag is used. RELATED TAGS If the verb form exists but simply does not agree with the subject, that considered is an agreement error (G_Agr)	wrong verb form (1 or more token)	GER: *Johannes [trinks] keine Milch. *Meine Eltern [bezahlenst] das Studium.; *Ich [kanne] nicht fahren; *... meine Rechte und Pflichten zu [weißen]; *Mein Mutte [wohn] noch in Vietnam. ITA: *E quando [vorebbi] {vorresti} il tuo vestito nuovo?; *Io [mangirò] {mangerò} poi una mela; *il sole [scalduva] {scaldava} molto forte; *puoi [venir] anche tu CZE: [radovám] {raduji} se. [pozváš] {pozveš} mě na oslavu. co si [přáte] {přeiete}	R
wrong inflection	case	G_Morphol_case_wrong	This tag is applied if the case of a substantive, adjective or pronoun is wrong. The form chosen exists, but it belongs to another case in the inflectional paradigm. The tag concerns all types of phrases containing substantives, pronouns and adjectives. Each wrong token in the phrase was tagged.	wrong token	GER: *Diese Erfahrung wurde [mich] späte gut helfen...; * Über eine positive Antwort wurde ich [mir] sehr freuen ...; *ich wollTe [dir] Fragen... ITA: *Lei (le) scrivo; *Come può arrivare a casa sua se te (ti) voglio venire a trovare CZE: čte romány a [chodí na procházce], chce [navštívit babičce]	R
wrong inflection	number	G_Morphol_numb_wrong	This tag is used if the number of a substantive, pronoun, article or adjective is wrong. The form chosen exists, but it belongs to another number in the inflectional paradigm. The error may occur in all types of phrases. Each wrong token in the phrase was tagged. LANGUAGE-SPECIFIC REMARKS ITA: Errors in the use of i/gli and il/lo are annotated with this tag, too.	wrong token	ITA: *L'autobus c'era solo 2 [volta]{volte}, *i miei [argomento] {argomenti}, *le [notizia attuale] {notizie attuali}, *restare qualche [giorni] {giorno}, *[Alle] {Alla} fine, *[la] {le} possibilità, *[i] {lo} sport, *[i] {gli} amici, * "Puoi chiamarla" instead of "Puoi chiamarle" GER: *Ich werde [zwei Woche] dort verbringen; *[zwei uralte Frage]; *[einige Postkarte] CZE: pracuje jako [knihovníci] {knihovnik}, vybíral si mezi [možností] {možnostmi}, znám tvoje [nápad] {nápady}	R
wrong inflection	gender	G_Morphol_gend_wrong	This tag is used if a substantive, pronoun or adjective or article contains a gender error. The form chosen exists, but it belongs to another gender in the inflectional paradigm. The error may occur in all types of phrases. Each wrong token in the phrase was tagged. RELATED TAGS For cases like *un'aiuta instead of "un aiuto", *passeggiati instead of "passeggiate" or *(ho) molto lavora instead of "(ho) molto lavoro" see discussion in G_Infect_noun .	wrong token	ITA: *Con riferimento [a] {alla} Sua lettera, * [I regoli] {le regole} di grammatica, *[Cara] {Caro} Giorgio, *[i miei congratulazioni] {le mie congratulazioni}, il [suo] {vostro} amico svizzero, *in [una] {un} ristorante. GER: *Ich brauche [eine große Wagen] für die Möbel.; *Es gibt [kein großes Unterschied]; *[Sprache und Kultur seines Heimats]; *ich wünsche dir [eine gute Wochenende] CZE: používám [tenhle tramvaj] {tuhle tramvaj}, neznám [žádnou takovou film] {žádný takový film}, to je už [můj druhý práce] {moje druhá práce}	R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
wrong inflection	ambiguous	G_Morphol_ambig_wrong	There are ambiguous cases of morphological errors for which it is very hard to decide if they are an error in case, gender, or number. The form exists in the inflectional paradigm of the word, but there is more than one "right" solution.	wrong token	ITA: *[il mio profession] {la mia professione}, *[ma queste notizia che cerchi è buono] GER: *[Nächste Monate] muss ich meine neue Arbeit beginnen (Numerus? Genus?); *Die Silvesternacht habe ich [mit meiner Kinder] verbracht.(Numerus? Kasus?); *um [einer Ort] zu besuchen... (Genus? Kasus?) CZE: vždy se těším na tvé dopis (number or gender?)	R
verb	tense	G_Verb_tns	If the learner chooses an erroneous tense, e.g. past tense instead of present tense (and vice versa), or present tense instead of future tense, this tag is used (see examples under a)). It also includes errors in the consecutio temporum (see examples under b)). RELATED TAGS The tag is not used for errors in verb formation (wrong person, number etc., see G_verb_compl) If the wrong form contains also errors in aspect, mood, voice or/and formation, several G_verb tags on the same token are applied.	wrong verbal form (1-3 tokens)	a)*včera [jdu] do kina, *zitra [studoval]; *gestern wir [kochen] gemeinsam, *nächste Woche [ging] er in den Urlaub; *ieri [va] al cinema, *la settimana prossima [sono stati] a Roma b) *Mi ha domandato se [ho] fretta {Mi ha domandato se avevo fretta}, *Carlo raccontò che Gianni non lo [chiamò] {Carlo raccontò che Gianni non lo aveva chiamato}, *Anna ha promesso che [passa] a prendermi {Anna ha promesso che sarebbe passata a prendermi}	R
verb	voice	G_Verb_vc	For confusions between the active and the passive voice, this tag is used. RELATED TAGS The tag is not used for errors in verb formation (wrong person, number etc., see G_verb_compl) If the wrong form contains also errors in aspect, mood, tense, or/and formation, several G_verb tags on the same token are used.	wrong verbal form (1-3 tokens)	*Jan [byl dán] mně tužku; *studenti [budou napsáni] test, *Peter [wurde gezeit] mir sein neues Buch, die Stadt [gründete] im Jahre 1234; *Giavanna mi [era data] la macchina, *il giorno [fissano] (era/è stato fissato), *nella mia città [viene cercato] personale	R
verb	mood	G_Verb_md	If a learner uses the imperative/conjunctive/indicative mood instead of the appropriate mood, this tag is used. RELATED TAGS The tag is not used for errors in verb formation (wrong person, number etc., see G_verb_compl) If the wrong form contains also errors in aspect, mood, voice, tense, or/and formation, several G_verb tags on the same token are used.	wrong verbal form (1-3 tokens)	*[Jdi] do města?, *[učila bych] to (instead of uč se to!); *er [würde gehen] gestern ins Kino (instead of ist gestern ins Kino gegangen/ging gestern...), *[Lerne] das? (instead of Lerne das!); *[Stai] bene!, Gianni [sarebbe andato] in città (instead of è andato in città)	R/U
verb	aspect	G_Verb_asp	This tag is applied for Czech and Italian. CZE: If the learner uses the imperfective instead of the perfective aspect, this tag is used. As the choice of aspect often depends on the subjective view on an action, only clear errors in use of aspect are annotated. If the context allows both aspectual forms, no annotation is done. ITA: If the learner uses the passato prossimo for a habitual action or the imperfetto for a finished, single action, this tag is used.	wrong verbal form (1-3 tokens)	CZE: pf. Instead of ipf.: *celý den [se naučil], ipf. Instead of pf. *[vařil] polévku a potom šel do kina ITA: imperfetto instead of pass.pross.: *sempre pensavo {ho sempre pensato} che voi due; *durante gli ultimi tempi aiutavo {ho aiutato}, *era {è stata} una esperienza molto interessante; pass.pross. instead of imperfetto: *l'acqua è stata {era} molto brutta, *non ho saputo {sapevo} cosa devo fare, *la stanza è stata {era} caldissima	R/U
verb	verb formation	G_Verb_compl_O G_Verb_compl_Ch G_Verb_compl_Ad G_Verb_compl_Pos	In this tag we distinguish errors in the formation of complex predicates (i.e. analytical verb forms, predicates with modals and copulative predicates). G_Verb_compl is used for: a) wrong form of the participle, wrong auxiliary (present instead of participle, infinitive instead of participle, "be" instead of "have" and viceversa, etc.) (G_Verb_compl_Ch) b) missing auxiliary / extra auxiliary (G_Verb_compl_O, G_Verb_compl_Ad) c) errors in infinitive forms (GER: includes missing "zu" of the "erweiterter Infinitiv mit 'zu'") (G_Verb_compl_Ch/O/Ad) d) missing or redundant copula (G_Verb_compl_O, G_Verb_compl_Ad) e) participle, auxiliary, copula in the wrong position (G_Verb_compl_Pos) f) problems in impersonal constructions (only for ITA and CZE) RELATED TAGS Wrong person or number are tagged as G_Agr. Wrong choice of modal verb is tagged as V_Sem_denot, wrong form of modal verb is G_Agr. Wrong form of copula is also G_Agr.	erroneous form in the verbal complex; missing verbal component --> existing one. missing copula -> whole clause.	a) *bude [dělá], *[učit] jsem se; *er wird [lese], *ich habe [lernen]; *Luigi ha [leggere] b) *kluci [jsou] běhali *my pracovali dlouho; *sie [gerannt], *sie [haben] arbeiteten; *Michele e Luca [andati] in città, *[ho] leggevo il libro, *[siamo] potuti giocare c) *musíš [přijdeš], *začal [pracuje]; *du musst [kommst]; *Mia mamma deve [viene] subito. * Ha cominciato di [lavora], *Diese zwei Frage richtig {zu} beantwortet ist nicht einfach. d) *Iva [studentka], *ve městě [má] knihovna, *Studentka [je] Iva přišla do školy; *Nico [Student], *Der Buchladen [hat] in der Stadt, *Die Studentin [ist] kam in die Schule; *Francesco [studente], *La libreria [ha] in città, *La studentessa [è] veniva a scuola. e) Dnes Karel milý je.f) f) * La città non se (si) vede ancora, *v Praze [on] stále se demonstuje	R/U
verb	main verb	G_Verb_main_O G_Verb_main_Ad	Tag used if in a learner's sentence there is a main verb missing (G_Verb_main_O) or there is a redundant main verb (G_Verb_main_Ad). RELATED TAGS If a copula is missing, G_Verb_Compil_O is used (for example: *Penso che questo compito {sia} abbastanza difficile.	redundant main verb In case of missing main verb ->whole clause	GER: *.. mit großem Interesse habe ich in XY Zeitung Ihre Anzeige {gelesen}; *Ich [nehme] besuche meine Tochter.; *Ich [wurd] wäre Ihnen sehr dankbar,.... CZE: *Chlapci si nejraději hrají s auty, dívky {zbožňují} panenky. [Jdu] pozvu Tě na párty. ITA: Missing: * Penso che questo compito {sia} abbastanza difficile; un anno fa {sono stata} in Germania per una settimana; {È} una domanda molto importante.	T/U

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
preposition		G_Prep_O G_Prep_Ad G_Prep_Ch G_Prep_Pos G_Prep_Merge G_Prep_Split	<p>This tag is applied if:</p> <p>a) a preposition is missing (G_Prep_O) b) a preposition is redundant (G_Prep_Ad) c) a preposition is incorrect (G_Prep_Ch). This, for CZE, includes missing parts of complex prepositions & errors in vocalization of prepositions d) a preposition is in the wrong position (G_Prep_Pos) V) a preposition is incorrectly contracted (G_Prep_Merge) VI) a preposition is incorrectly split (G_Prep_Split)</p> <p>LANGUAGE-SPECIFIC REMARK FOR ITALIAN G_Prep_Ch is not applied if a preposition with an article is used instead of a simple one and vice versa (i.e. *Giocare al calcio -> "al" instead of "a", "su" instead of "sulla"). Here, article tags are used (G_Art_Ad; G_Art_O). In the same "componential view", if an articulated preposition is wrongly added (i.e. *Mi piacerebbe frequentare sul questo corso), we use both tags G_Prep_ad and G_Art_Ad.</p> <p>RELATED TAGS If a preposition is incorrectly contracted with or split from an article, the tags G_Art_Merge or G_Art_Split are used. Errors in the use of prepositions often lead to wrong formulaic sequences. For deciding whether to tag a G_Art or a V_form_FS, for each case it was decided whether that form could also grammatically exist in another context.</p>	1 token; in case of missing prep. --> phrase	<p>a) *myslím [maminku] (na maminku), *dívá se [televizi]; *ich warte [Antwort], *dann warte ich [dich]; *mi da [possibilità] girare e conoscere, *una sera la [settimana] b) *bojí se [od] psa, *věří [k] učitelce, *kannst du [bei] mir helfen?, *er ist [in] zur Schule gegangen; *è stato un disastro [ad] andare al lago, *non è facile per me [di] viaggiare c) *sedí si [do] židli, *jdu [do] doktora, *[v] škole; *ich soll [am] 8. Januar bis 12. Januar besuchen, *viel Spaß [zum] deine Arbeit; *assistente [alla] sua azienda, *un giorno [nel] posto di lavoro d) přijít hodinou před. {přijít před hodinou}, *Er ist gekommen eine Stunde vor; *È dallo caduto albero {è caduto dall'albero} e) předemnou {přede mnou}, vormir {vor mir}; *perme f) okol o, *her um {herum}</p>	R/U
article		G_Art_O G_Art_Ch G_Art_Ad G_Art_Merge G_Art_Split	<p>This tag is used in the following cases:</p> <p>a) missing article (G_Art_O) b) wrong article (confusion between definite and indefinite article or other article words, confusion between forms of the same gender & number (ITA: il vs. lo; i vs. gli) (G_Art_Ch) c) article is redundant (G_Art_Ad) d) preposition and article are wrongly contracted (G_Art_Merge) e) preposition and article are not contracted (G_Art_Split)</p> <p>LANGUAGE-SPECIFIC REMARK GER/ITA: For deciding if merging or splitting a preposition and an article is an error, please the rules in canoo were followed (www. canoo.net). G_Art_Merge is used only if merging is not possible, and G_Art_Split is used only if merging is obligatory.</p> <p>RELATED TAGS If an article is involved, G_Art_Merge and G_Art_Split are used, not G_Prep_Merge.</p>	1 or 2 tokens (1 token: wrong article, 2 tokens: prep + art not contracted) missing article: phrase	<p>a) *habe [litauische Staatsangehörigkeit], *[in] folgenden Bericht, *[mia insegnante] si chiama, *sono [persona] molto socievole b) *[il] mese fa siamo andati, *c'è [una] possibilità; *sono venuto con [dei] amici; *vorrei del zucchero; *ich bringe [etwas] GeschÄ*nk..., *[il] studente, *[i] amici, *dedicandomi [ai] altri c) *[Die] letzte Woche habe ich eine neue Arbeit gefunden, *[zum] euch, *[vom] euch, *[zum] der] Dorf; *con [la] mia moglie; *giocare [al] calcio d) *[per]le altre {per le}, *[Fralle] {fra le} e) *[su l']erba {sull'}, *da la {dalla}</p>	T
conjunction		G_Conj_O G_Conj_Ad G_Conj_Ch G_Conj_Pos	<p>This tag is applied if:</p> <p>a) a conjunction is missing (for "normal" conjunctions, relative pronouns, and CZE: conditional conjunction (aby)) (G_Conj_O) b) a conjunction is redundant (G_Conj_Ad) c) a conjunction is incorrect (choice) (G_Conj_Ch) d) a conjunction is in the wrong position (G_Conj_Pos)</p> <p>(b) and c) include relative pronouns</p>	1 token; if conjunction is missing --> clause supposed to contain the missing element	<p>a) *řekl mu Petr to nezná, *ich möchte dir sagen, ich bin glücklich {dass ich glücklich bin}, *speriamo ti stai bene//*viděla muže Jana neznala; *er füttert den Hund, nicht ihm gehört; *Io studente, sta in classe {,che sta...}//*řekl mu nedělal to {řekl mi, aby to nedělal} b) *doporučil mi [aby] navštívit doktora, *bál se, [že] jít k doktorovi; *er half mir [dass] ich aufstehe, *Karl kam [um] [für] helfen; *capisco [che] il problema, *da due anni [che] sto imparando inglese c) *řekl, [protože] má čas, *přijedeme, [kdy] nebude pršet, *vzal kabát, [kdo] ležel na židli, *[abych] měl peníze, koupil bych si auto, *dal si kávu, [kterého] má rád; *sar à meglio [quando] uno di voi mi chiama d) nepíšeš, nevádí mi to ale</p>	R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
clitic		G_Clit_Ch G_Clit_O G_Clit_Ad	<p>This tag is language-specific (ITA) and is to be used in the following cases:</p> <p>a) wrong person/case (G_Clit_Ch) b) missing clitic (G_Clit_O) c) missing pronoun particle/pronoun particle+verb (G_Clit_O) d) pronoun particle not required (G_Clit_Ad)</p> <p>RELATED TAGS If the omitted clitic is a missing obligatory object, G_Clit_O AND G_Valency_complnumb_O are used, e.g. in *Ho dimenticato di scrivere [scriverlo] prima. If a reference problem is caused by a missing/wrong clitic (e.g. "vederlo" instead of "vederti"), a C_coh_ref tag is added.</p>	a), b), d), token with the wrong, added or wrongly placed element c) verb	<p>a) puoi [chiamarla] {puoi chiamarmi} ; vorrei [mandarla] il mio curriculum {vorrei mandarLe il...}</p> <p>b) ho dimenticato di [scrivere] prima {ho dimenticato di scriverlo prima}</p> <p>c) [sono] le lezioni {ci sono le lezioni}; non problema {non c'è problema}</p> <p>d) il lago non [c'era] vicino all'appartamento {il lago non era vicino...}</p>	T
part of speech		G_POS	<p>This tag is supposed to identify the erroneous use of parts of speech (word classes). If a learner uses a noun instead of a verb, or an adjective instead of an adverb, for example, this tag is used.</p>	one token (wrong POS)	<p>ITA: *Tra cui (queste) l'inglese è la più facile; *Ho dimenticato di (come) parlare tedesco; *Ogni (I) bambini dovrebbero studiare; *Una signora molta (molto) simpatica qui (che) si chiama Maria; *Non ha sentito buono (bene); *Quando (Quanto) costa il corso?</p> <p>GER: *Ich freue mich für unsere [besucht] {Besuch}; *Ich bin sehr flexibel und [Mobilität] {mobil}; *Kannst du mich [Hilfe] {helfen}; *[Zusammenfassung] lässt sich sagen {zusammenfassend}</p> <p>CZE: *bylo to pro mě [potěšený] {potěšení}, *musíš myslet [rozumný] {rozumně}, zdravý život bez [nemocni] {nemoci}, doufám, že počasí bude [dobře] {dobře}</p>	R

MERLIN ANNOTATION SCHEME (I)

GRAMMAR TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
word order	word order in main clause	G_Wo_womaincl	This tag is used if an error occurs in a main clause in the following cases: a) verb in the wrong position b) wrong order of syntactical items such as attribute, subject, object. In some cases, both the position of the verb AND the position of subject, object etc. are wrong. No differentiation is made in this case. The tag also applies to word order errors in questions and exclamations. LANGUAGE-SPECIFIC REMARKS GER: includes errors in the use of "zu" (only wrong position, not missing!) CZE: includes errors in the use of "se"/"si" (only wrong position, not missing!) RELATED TAGS If a participle, auxiliary, or copula is in the wrong position, it is tagged as G_Verb_compl_Pos.	wrong clause including punctuation mark	a) *[Včera učil jsem se česky.], *[Vielleicht du könntest mir bei meine Wohnungssuche helfen.], *[Sollst du Wasser und Bikini mitbringen.]; *[Per trovarlo ti vorrei aiutare.] b) *[Známe film český.], *[Učil se jsem česky.], *[Könntest du bitte mich anrufen?], *[Köntest Du bitte mir helfen bei der Wohnungssuche?]; *[Mia famiglia sta anche bene.], *[Vorrei andare a ti trovare.]	R/T/U
word order	word order in subordinate clause	G_Wo_wosubcl	This tag is used if an error occurs in a subordinate clause in the following cases: a) verb in the wrong position b) wrong order of syntactical items such as attribute, subject, object. In some cases, both the position of the verb AND the position of subject, object etc. are wrong. No differentiation is made in this case. The tag also applies to word order errors in questions and exclamations.	wrong clause including punctuation mark	a) První je těžký znát, [co učil se.], *Ich wollte wissen, [welche Sprache wird in der Familie gesprochen.], *wenn haben Sie Zeit.], dann bitte sagen Sie mir. *Gerne würde ich auch wissen [welche Aufenthaltsformalitäten verlangen Sie von mir?] b) Maminka myslí, [že musí pracovat moc pěkně pro lepší budoucnost.], *Ich habe nicht verstanden, [von welche fährt der Zug Bahnhof.]	R/T/U
negation	general negation error	G_Neg_neggen_Pos G_Neg_neggen_O G_Neg_neggen_Ch G_Neg_neggen_Ad	This tag is used in case of a) wrong position of negation expressions (G_Neg_neggen_Pos) b) missing part of negation (G_Neg_neggen_O) c) wrong use of negation words (e.g., GER: nicht, nein, kein; ITA: no, non; CZE: ne, žádný (G_Neg_neggen_Ch) d) redundant/wrongly added negation word (G_Neg_neggen_Ad)	a) _pos: 1 token; 2 tokens for CZE (neg.word and verb, if neg. word wrongly distributed) b) _o: 1 or more token - the POS to be negated; c) _ch: 1 token - the negation word	a) *[mám ne] kávu, *[půjdu neráno], *Ich habe Hunger [keinen]; *Io credo [non]. b) *On [ne] velký; • [Io mangio nè carne nè pesce] {Io non mangio né carne né pesce}; *Luca non va a scuola perché ne ha voglia; {Luca non va a scuola perché non ne ha voglia} *Non è né chiaro scuro] {Non è né chiaro né scuro}. c) *Bohužel, nemám [ne] čas; *Ich habe [nicht] Zeit. *Er wird dort arbeiten [nein]; *[Non], viene più tardi. d) *Man kann nicht auf solchen grössen Teil seiner Persönlichkeit zu verzichten, ohne seine psychische Gesundheit [nicht] zu schaden.	R
negation	double negation	G_Neg_negdoub	This is a single-language tag that is only applied for CZECH. In Czech, all negated pronouns require a negated verb form. This tag refers to the missing negation particle "ne" at the verb and vice versa.	1 token (verb, existing part of the double negation)	*[mám] žádný čas {nemám žádný čas} *nikdo [volal] → {nikdo nevolal} *nemám [nějaký] čas {nemám žádný čas }	R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
verb valency (obligatory arguments)	complement number	G_Valency_complnumb_O G_Valency_complnumb_Ad	In Merlin, the erroneous <i>presence or absence</i> of obligatory arguments is annotated. Free complements (Helbig: "freie Angaben") are not tagged. The tags are specified as follows: a) omission of a subject or obligatory object (G_Valency_complnumb_O) b) wrong addition of a subject or obligatory object (G_Valency_complnumb_Ad). This tag was used as scarcely as possible. Annotators were asked to rely on their native speaker competence to decide whether an explicit subject really is an error (annotated) or a question of style (not annotated). Of course, obligatory arguments may be missing because the learner gives the impression to have created an elliptic construction consciously. In those cases, the missing argument is not tagged. A redundant subject may occur in Italian and Czech. Often, the use of an explicit subject in these languages is subjective and does not necessarily represent an error. The tag was used only in the case of an obvious error clearly contradicting target language usage norms. LANGUAGE-SPECIFIC REMARKS GER: "Es" as correlate to an obligatory or facultative complement may be wrongly added or absent. For both, the G_Valency_complnumb tag is used. RELATED TAGS The omission of a subject or obligatory object usually leads to a reference problem. The coherence/cohesion tag (C_coh_ref) was then used additionally. If a learner omits a clitic that is an obligatory object, G_Clit_O is used, too, e.g. *Ho dimenticato di scrivere [scriverlo] prima.	wrong argument (usually 1 token) or whole clause including punctuation mark (if argument is missing)	a) *Já vstávám v 5, [ale vstáváš v 8.] {...ale ty vstáváš v 8.} *Ich liebe {dich}. *Er hat uns nicht gesagt, ob {er} kommen will. *[Spero che possa aiutarLa.]. {Spero che questo possa aiutarLa.}, *[Forse con cui si parla in chat non è quella persona che tu pensi] {Forse la persona con cui si parla in chat non è quella persona che tu pensi}." b) *Petr vstává v 6 hodin. On nesnídá, protože [on] nemá hlad.; *Susanne denkt [sie] zu viel nach. *Che lavoro [tu] fai adesso? *Secondo me [tu] hai lavorato. *Che cosa di preciso [tu] vuoi fare?, * Adesso [la mia professione] sono insegnante.	R
agreement (subject and verb)		G_Agr	An agreement error in MERLIN includes wrong grammatical forms in the following combinations of subject and verb (incl. copulative predicates): a) wrong form (person, number) of the verb b) wrong form (person, number) of an auxiliary or participle c) wrong part of speech (POS) in copulative predicate (infinitive instead of adjective etc.) d) wrong form (person, number) of copula Of course, combinations of adjective and substantive, pronoun and substantive, numerals and substantives etc. may contain agreement errors, too. In Merlin, though, it was decided to annotate these as morphological errors (see G_Morphol tags). This way, the difficult categorization of a phenomenon as either an agreement error or a complement error was avoided. RELATED TAGS If the subject is not explicit, but inferable from the context, and the verb form is grammatically correct, no G_Agr tag is applied. The phenomenon is tagged as a coherence problem (TH2 & EA2, C_coh_ref). Examples: *Secondo che tipo lavoro vuole fare? *Come stai? Che cosa fai? Ha anche molto lavoro? (In both cases the subject suggested by the context is "tu" and the speaker was addressing it informally in the rest of the text).	wrong form (usually 1 token)	I) a) *Jana [čtu], *Petr [nerozumět]; *chci se [učí]; *Jana [lese], *Peter [verstehen]; *Gianni [leggo], *Chiara [capire]; *Noi [volevano] un gelato; *Tu [sapeva] questo. b) *Jana [jsi] udělala, *přišel [jsme]; *[abychom] mohl {abych mohl}; *Jana [has] gelesen, *wie [seid] gekommen; *Gianni [abbiamo] letto, *(noi) [sono] venuti, *Siamo potuti (abbiamo potuto) giocare c) *Petr je [unavit]; *Jana ist [müde werden]; *Chiara è [stancare] d) *Petr [jsou] unavený; *Jana [sind] müde; *Giovanni e Luca [siamo] stanchi	R/C
reflexivity	reflexive pronoun	G_Refl_pronrefl_O G_Refl_pronrefl_Ad G_Refl_pronrefl_Ch G_Refl_pronrefl_Pos	Errors in the use of reflexive pronouns in MERLIN are tagged in case of: a) missing pronoun (G_Refl_pronrefl_O) b) wrongly added reflexive pronoun (G_Refl_pronrefl_Ad) c) wrong choice of reflexive pronoun (also in the object position) or using personal pronoun instead of the reflexive one (G_Refl_pronrefl_Ch) d) wrong position of the reflexive pronoun (G_Refl_pronrefl_Pos) RELATED TAGS In case of errors in reflexives in the object position, it is hard to decide whether it is an agreement error (G_Agr) or an error in choice of the right pronoun. In this cases we tag the error as reflexivity error.	wrongly positioned reflexive pronoun or token containing the reflexive pronoun, in case of missing reflexive: verb	a) *[dívám] na televizi, *ráno [maluju] {instead of: ráno se maluju}, *er [entschuldigt], *Laura und Ferdinand reden [sich]; *Michele [rivolge] à sua sorella b) *dnes [si] neučí, *vidím [se] Petra; *[mi] desidero una camera singola c) *myju [se] ruce, *myslím [se] to, *smála [si]; *abychom [nás] mohli vidět abychom {se} mohli vidět; *[se] {si} lava ogni mattina, *[si] {ci} rivolgiamo a Lei d) Já těším na Tvou oslavu. Chtěla bych zeptat se., *[mi] non posso incontrare con tu	R/T
reflexivity	possessive reflexive pronoun	G_Refl_pronreflposs	This is a single-language tag that is only applied for CZECH. Czech distinguishes between non-reflexive possessive pronouns (můj, tvůj...) and reflexive possessive pronouns (svůj, svoje...). Confusion between these are annotated with G_Refl_pronreflposs RELATED TAGS If there is also a form problem (*stojí před její dům instead of *svým domem), the pronoun will be tagged as agreement error (G_Agr) and as G_Refl_pronreflposs	wrong reflexive pronoun	*potřebuju [moji] knihu, * vidím [mého] otce	R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
inexistent inflection	adjective inflection	G_Infect_adj_inexist	This tag is used if an adjective form contains errors, i.e. if the learner uses an inflected form of the adjective that does not exist in the inflectional paradigm at all. The tag includes wrong forms of comparative/ superlative, too. LANGUAGE-SPECIFIC REMARK GER: Beyond that, the tag is used if the learner uses the determinate paradigm instead of the indeterminate. In this case, the form may exist but it belongs to another paradigm. RELATED TAGS If the form exists but belongs to another case or number, the G_Morphol tags are used.	wrong adjective form (1 or more token)	GER: *Das ist ein schöne Haus. (determinate instead of indeterminate) *ein [blaus] Himmel {blauer}; *... mit [freundlich] Grüßen ITA: *una persona [pedanta] {pedante}, *un tempo [mutevolo] {mutevole}; *polli [arrosti] {polli arrosto}; *poltrone [lille] {poltrone lilla} CZE: v [následujícím] {následujícím} měsíci, s [cizou] {cizi} nabídkou	R
inexistent inflection	noun inflection	G_Infect_noun_inexist	Generally, substantives, articles, adjectives, numerals, and pronouns count as nouns. In Merlin, we decided to separate errors in the inflection of adjectives from all other types of nouns. Therefore, this tag is used if the form of a substantive, article, numeral or pronoun contains an error. It only applies if the learner uses an inflected form that does not exist in the inflectional paradigm at all. RELATED TAGS If the form exists but simply belongs to another case or number, the G_Morphol tags are used. If a learner writes, e.g. *un'aiuta instead of "un aiuto" or *passeggiati instead of "passeggiate", we tag G_Infect_noun and not V_word/fs_nonexist . The reason is that we assume that the main problem is not that the word does not exist but that the learner has chosen a wrong gender which is a form that does not exist in the paradigm of the noun. The same applies for with G_Infect_verb and G_Infect_adj . Furthermore, in cases like *un'aiuta vs "un aiuto", neither the article (G_Art) nor the apostrophe error (O_Apostr) are tagged. These errors depend on the gender error of the noun, while the learner was consistent in the (erroneous) form used.	wrong noun form (1 token)	GER: *[euche] [Fahrrade], *er gab es [mire] {mir}; * welche Rechte und [Pfichte]... ITA: *Con [personi] {persone}, *chiedere [un'aiuta] {un aiuto}, *[un Suo Risposto] {una Sua risposta} CZE: dej to [mojemu] {mému} kamarádovi, promluvit k [lidům] {lidem}, o [ticicích] {tisících} nápady	R
inexistent inflection	verb inflection	G_Infect_verb_inexist	If the learner uses a verb form that does not exist in the paradigm of the verb, this tag is used. RELATED TAGS If the verb form exists but simply does not agree with the subject, that considered is an agreement error (G_Agr)	wrong verb form (1 or more token)	GER: *Johannes [trinks] keine Milch. *Meine Eltern [bezahlenst] das Studium.; *Ich [kanne] nicht fahren; *... meine Rechte und Pflichten zu [weißen]; *Mein Mutte [wohn] noch in Vietnam. ITA: *E quando [vorebbi] {vorresti} il tuo vestito nuovo?; *Io [mangirò] {mangerò} poi una mela; *il sole [scalduva] {scaldava} molto forte; *puoi [venir] anche tu CZE: [radovám] {raduji} se. [pozváš] {pozveš} mě na oslavu. co si [přáte] {přeiete}	R
wrong inflection	case	G_Morphol_case_wrong	This tag is applied if the case of a substantive, adjective or pronoun is wrong. The form chosen exists, but it belongs to another case in the inflectional paradigm. The tag concerns all types of phrases containing substantives, pronouns and adjectives. Each wrong token in the phrase was tagged.	wrong token	GER: *Diese Erfahrung wurde [mich] späte gut helfen...; * Über eine positive Antwort wurde ich [mir] sehr freuen ...; *ich wollTe [dir] Fragen... ITA: *Lei (le) scrivo; *Come può arrivare a casa sua se te (ti) voglio venire a trovare CZE: čte romány a [chodí na procházce], chce [navštívit babičce]	R
wrong inflection	number	G_Morphol_numb_wrong	This tag is used if the number of a substantive, pronoun, article or adjective is wrong. The form chosen exists, but it belongs to another number in the inflectional paradigm. The error may occur in all types of phrases. Each wrong token in the phrase was tagged. LANGUAGE-SPECIFIC REMARKS ITA: Errors in the use of i/gli and il/lo are annotated with this tag, too.	wrong token	ITA: *L'autobus c'era solo 2 [volta]{volte}, *i miei [argomento] {argomenti}, *le [notizia attuale] {notizie attuali}, *restare qualche [giorni] {giorno}, *[Alle] {Alla} fine, *[la] {le} possibilità, *[i] {lo} sport, *[i] {gli} amici, * "Puoi chiamarla" instead of "Puoi chiamarle" GER: *Ich werde [zwei Woche] dort verbringen; *[zwei uralte Frage]; *[einige Postkarte] CZE: pracuje jako [knihovníci] {knihovnik}, vybíral si mezi [možností] {možnostmi}, znám tvoje [nápad] {nápady}	R
wrong inflection	gender	G_Morphol_gend_wrong	This tag is used if a substantive, pronoun or adjective or article contains a gender error. The form chosen exists, but it belongs to another gender in the inflectional paradigm. The error may occur in all types of phrases. Each wrong token in the phrase was tagged. RELATED TAGS For cases like *un'aiuta instead of "un aiuto", *passeggiati instead of "passeggiate" or *(ho) molto lavora instead of "(ho) molto lavoro" see discussion in G_Infect_noun .	wrong token	ITA: *Con riferimento [a] {alla} Sua lettera, * [I regoli] {le regole} di grammatica, *[Cara] {Caro} Giorgio, *[i miei congratulazioni] {le mie congratulazioni}, il [suo] {vostro} amico svizzero, *in [una] {un} ristorante. GER: *Ich brauche [eine große Wagen] für die Möbel.; *Es gibt [kein großes Unterschied]; *[Sprache und Kultur seines Heimats]; *ich wünsche dir [eine gute Wochenende] CZE: používám [tenhle tramvaj] {tuhle tramvaj}, neznám [žádnou takovou film] {žádný takový film}, to je už [můj druhý práce] {moje druhá práce}	R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
wrong inflection	ambiguous	G_Morphol_ambig_wrong	There are ambiguous cases of morphological errors for which it is very hard to decide if they are an error in case, gender, or number. The form exists in the inflectional paradigm of the word, but there is more than one "right" solution.	wrong token	ITA: *[il mio profession] {la mia professione}, *[ma queste notizia che cerchi è buono] GER: *[Nächste Monate] muss ich meine neue Arbeit beginnen (Numerus? Genus?); *Die Silvesternacht habe ich [mit meiner Kinder] verbracht.(Numerus? Kasus?); *um [einer Ort] zu besuchen... (Genus? Kasus?) CZE: vždy se těším na tvé dopis (number or gender?)	R
verb	tense	G_Verb_tns	If the learner chooses an erroneous tense, e.g. past tense instead of present tense (and vice versa), or present tense instead of future tense, this tag is used (see examples under a)). It also includes errors in the consecutio temporum (see examples under b)). RELATED TAGS The tag is not used for errors in verb formation (wrong person, number etc., see G_verb_compl) If the wrong form contains also errors in aspect, mood, voice or/and formation, several G_verb tags on the same token are applied.	wrong verbal form (1-3 tokens)	a)*včera [jdu] do kina, *zitra [studoval]; *gestern wir [kochen] gemeinsam, *nächste Woche [ging] er in den Urlaub; *ieri [va] al cinema, *la settimana prossima [sono stati] a Roma b) *Mi ha domandato se [ho] fretta {Mi ha domandato se avevo fretta}, *Carlo raccontò che Gianni non lo [chiamò] {Carlo raccontò che Gianni non lo aveva chiamato}, *Anna ha promesso che [passa] a prendermi {Anna ha promesso che sarebbe passata a prendermi}	R
verb	voice	G_Verb_vc	For confusions between the active and the passive voice, this tag is used. RELATED TAGS The tag is not used for errors in verb formation (wrong person, number etc., see G_verb_compl) If the wrong form contains also errors in aspect, mood, tense, or/and formation, several G_verb tags on the same token are used.	wrong verbal form (1-3 tokens)	*Jan [byl dán] mně tužku; *studenti [budou napsáni] test, *Peter [wurde gezeit] mir sein neues Buch, die Stadt [gründete] im Jahre 1234; *Giavanna mi [era data] la macchina, *il giorno [fissano] (era/è stato fissato), *nella mia città [viene cercato] personale	R
verb	mood	G_Verb_md	If a learner uses the imperative/conjunctive/indicative mood instead of the appropriate mood, this tag is used. RELATED TAGS The tag is not used for errors in verb formation (wrong person, number etc., see G_verb_compl) If the wrong form contains also errors in aspect, mood, tense, or/and formation, several G_verb tags on the same token are used.	wrong verbal form (1-3 tokens)	*[Jdi] do města?, *[učila bych] to (instead of uč se to!); *er [würde gehen] gestern ins Kino (instead of ist gestern ins Kino gegangen/ging gestern...), *[Lerne] das? (instead of Lerne das!); *[Stai] bene!, Gianni [sarebbe andato] in città (instead of è andato in città)	R/U
verb	aspect	G_Verb_asp	This tag is applied for Czech and Italian. CZE: If the learner uses the imperfective instead of the perfective aspect, this tag is used. As the choice of aspect often depends on the subjective view on an action, only clear errors in use of aspect are annotated. If the context allows both aspectual forms, no annotation is done. ITA: If the learner uses the passato prossimo for a habitual action or the imperfetto for a finished, single action, this tag is used.	wrong verbal form (1-3 tokens)	CZE: pf. Instead of ipf.: *celý den [se naučil], ipf. Instead of pf. *[vařil] polévku a potom šel do kina ITA: imperfetto instead of pass.pross.: *sempre pensavo {ho sempre pensato} che voi due; *durante gli ultimi tempi aiutavo {ho aiutato}, *era {è stata} una esperienza molto interessante; pass.pross. instead of imperfetto: *l'acqua è stata {era} molto brutta, *non ho saputo {sapevo} cosa devo fare, *la stanza è stata {era} caldissima	R/U
verb	verb formation	G_Verb_compl_O G_Verb_compl_Ch G_Verb_compl_Ad G_Verb_compl_Pos	In this tag we distinguish errors in the formation of complex predicates (i.e. analytical verb forms, predicates with modals and copulative predicates). G_Verb_compl is used for: a) wrong form of the participle, wrong auxiliary (present instead of participle, infinitive instead of participle, "be" instead of "have" and viceversa, etc.) (G_Verb_compl_Ch) b) missing auxiliary / extra auxiliary (G_Verb_compl_O, G_Verb_compl_Ad) c) errors in infinitive forms (GER: includes missing "zu" of the "erweiterter Infinitiv mit 'zu'") (G_Verb_compl_Ch/O/Ad) d) missing or redundant copula (G_Verb_compl_O, G_Verb_compl_Ad) e) participle, auxiliary, copula in the wrong position (G_Verb_compl_Pos) f) problems in impersonal constructions (only for ITA and CZE) RELATED TAGS Wrong person or number are tagged as G_Agr. Wrong choice of modal verb is tagged as V_Sem_denot, wrong form of modal verb is G_Agr. Wrong form of copula is also G_Agr.	erroneous form in the verbal complex; missing verbal component --> existing one. missing copula -> whole clause.	a) *bude [dělá], *[učit] jsem se; *er wird [lese], *ich habe [lernen]; *Luigi ha [leggere] b) *kluci [jsou] běhali *my pracovali dlouho; *sie [gerannt], *sie [haben] arbeiteten; *Michele e Luca [andati] in città, *[ho] leggevo il libro, *[siamo] potuti giocare c) *musíš [přijdeš], *začal [pracuje]; *du musst [kommst]; *Mia mamma deve [viene] subito. * Ha cominciato di [lavora], *Diese zwei Frage richtig {zu} beantwortet ist nicht einfach. d) *Iva [studentka], *ve městě [má] knihovna, *Studentka [je] Iva přišla do školy; *Nico [Student], *Der Buchladen [hat] in der Stadt, *Die Studentin [ist] kam in die Schule; *Francesco [studente], *La libreria [ha] in città, *La studentessa [è] veniva a scuola. e) Dnes Karel milý je.f) f) * La città non se (si) vede ancora, *v Praze [on] stále se demonstuje	R/U
verb	main verb	G_Verb_main_O G_Verb_main_Ad	Tag used if in a learner's sentence there is a main verb missing (G_Verb_main_O) or there is a redundant main verb (G_Verb_main_Ad). RELATED TAGS If a copula is missing, G_Verb_Compil_O is used (for example: *Penso che questo compito {sia} abbastanza difficile.	redundant main verb In case of missing main verb ->whole clause	GER: *... mit großem Interesse habe ich in XY Zeitung Ihre Anzeige {gelesen}; *Ich [nehme] besuche meine Tochter.; *Ich [wurd] wäre Ihnen sehr dankbar,.... CZE: *Chlapci si nejraději hrají s auty, dívky {zbožňují} panenky. [Jdu] pozvu Tě na párty. ITA: Missing: * Penso che questo compito {sia} abbastanza difficile; un anno fa {sono stata} in	T/U

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
preposition		G_Prep_O G_Prep_Ad G_Prep_Ch G_Prep_Pos G_Prep_Merge G_Prep_Split	<p>This tag is applied if:</p> <p>a) a preposition is missing (G_Prep_O)</p> <p>b) a preposition is redundant (G_Prep_Ad)</p> <p>c) a preposition is incorrect (G_Prep_Ch). This, for CZE, includes missing parts of complex prepositions & errors in vocalization of prepositions</p> <p>d) a preposition is in the wrong position (G_Prep_Pos)</p> <p>V) a preposition is incorrectly contracted (G_Prep_Merge)</p> <p>VI) a preposition is incorrectly split (G_Prep_Split)</p> <p>LANGUAGE-SPECIFIC REMARK FOR ITALIAN G_Prep_Ch is not applied if a preposition with an article is used instead of a simple one and vice versa (i.e. *Giocare al calcio -> "al" instead of "a", "su" instead of "sulla"). Here, article tags are used (G_Art_Ad; G_Art_O). In the same "componential view", if an articulated preposition is wrongly added (i.e. *Mi piacerebbe frequentare sul questo corso), we use both tags G_Prep_ad and G_Art_Ad.</p> <p>RELATED TAGS If a preposition is incorrectly contracted with or split from an article, the tags G_Art_Merge or G_Art_Split are used. Errors in the use of prepositions often lead to wrong formulaic sequences. For deciding whether to tag a G_Art or a V_form_FS, for each case it was decided whether that form could also</p>	1 token; in case of missing prep. --> phrase	<p>a) *myslím [maminku] (na maminku), *dívá se [televizi]; *ich warte [Antwort], *dann warte ich [dich]; *mi da [possibilità] girare e conoscere, *una sera la [settimana]</p> <p>b) *bojí se [od] psa, *věří [k] učitelce, *kannst du [bei] mir helfen?, *er ist [in] zur Schule gegangen; *è stato un disastro [ad] andare al lago, *non è facile per me [di] viaggiare</p> <p>c) *sedí si [do] židli, *jdu [do] doktora, *[v] škole; *ich soll [am] 8. Januar bis 12. Januar besuchen, *viel Spaß [zum] deine Arbeit; *assistente [alla] sua azienda, *un giorno [nel] posto di lavoro</p> <p>d) přijít hodinou před. {přijít před hodinou}, *Er ist gekommen eine Stunde vor; *È dallo caduto albero {è caduto dall'albero}</p> <p>e) předemnou {přede mnou}, vormir {vor mir}; *perme</p> <p>f) okol o, *her um {herum}</p>	R/U
article		G_Art_O G_Art_Ch G_Art_Ad G_Art_Merge G_Art_Split	<p>This tag is used in the following cases:</p> <p>a) missing article (G_Art_O)</p> <p>b) wrong article (confusion between definite and indefinite article or other article words, confusion between forms of the same gender & number (ITA: il vs. lo; i vs. gli) (G_Art_Ch)</p> <p>c) article is redundant (G_Art_Ad)</p> <p>d) preposition and article are wrongly contracted (G_Art_Merge)</p> <p>e) preposition and article are not contracted (G_Art_Split)</p> <p>LANGUAGE-SPECIFIC REMARK GER/ITA: For deciding if merging or splitting a preposition and an article is an error, please the rules in canoo were followed (www. canoo.net). G_Art_Merge is used only if merging is not possible, and G_Art_Split is used only if merging is obligatory.</p> <p>RELATED TAGS If an article is involved, G_Art_Merge and G_Art_Split are used, not G_Prep_Merge.</p>	1 or 2 tokens (1 token: wrong article, 2 tokens: prep + art not contracted) missing article: phrase	<p>a) *habe [litauische Staatsangehörigkeit], *[in] folgenden Bericht, *[mia insegnante] si chiama, *sono [persona] molto socievole</p> <p>b) *[il] mese fa siamo andati, *c'è [una] possibilità; *sono venuto con [dei] amici; *vorrei del zucchero; *ich bringe [etwas] GeschÃnk..., *[il] studente, *[i] amici, *dedicandomi [ai] altri</p> <p>c) *[Die] letzte Woche habe ich eine neue Arbeit gefunden, *[zum] euch, *[vom] euch, *[zum der] Dorf; *con [la] mia moglie; *giocare [al] calcio</p> <p>d) *[per]le altre {per le}, *[Fr]alle {fra le}</p> <p>e) *[su l']erba {sull'}, *da la {dalla}</p>	T
conjunction		G_Conj_O G_Conj_Ad G_Conj_Ch G_Conj_Pos	<p>This tag is applied if:</p> <p>a) a conjunction is missing (for "normal" conjunctions, relative pronouns, and CZE: conditional conjunction (aby)) (G_Conj_O)</p> <p>b) a conjunction is redundant (G_Conj_Ad)</p> <p>c) a conjunction is incorrect (choice) (G_Conj_Ch)</p> <p>d) a conjunction is in the wrong position (G_Conj_Pos)</p> <p>(b) and c) include relative pronouns)</p>	1 token; if conjunction is missing --> clause supposed to contain the missing element	<p>a) *řekl mu Petr to nezná, *ich möchte dir sagen, ich bin glücklich {dass ich glücklich bin}, *speriamo ti stai bene/**viděla muže Jana neznala; *er füttert den Hund, nicht ihm gehört; *Io studente, sta in classe {,che sta...}/**řekl mu nedělal to {řekl mu, aby to nedělal}</p> <p>b) *doporučil mi [aby] navštívit doktora, *bál se, [že] jít k doktorovi; *er half mir {dass} ich aufstehe, *Karl kam [um] [für] helfen; *capisco {che} il problema, *da due anni {che} sto imparando inglese</p> <p>c) *řekli, {protože} má čas, *přijdeme, [kdy] nebude pršet, *vzal kabát, [kdo] ležel na židli, *[abych] měl peníze, koupil bych si auto, *dal si kávu, [kterého] má rád; *sar à meglio [quando] uno di voi mi chiama</p> <p>d) nepíšeš, nevádí mi to ale</p>	R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
clitic		G_Clit_Ch G_Clit_O G_Clit_Ad	<p>This tag is language-specific (ITA) and is to be used in the following cases:</p> <p>a) wrong person/case (G_Clit_Ch) b) missing clitic (G_Clit_O) c) missing pronoun particle/pronoun particle+verb (G_Clit_O) d) pronoun particle not required (G_Clit_Ad)</p> <p>RELATED TAGS If the omitted clitic is a missing obligatory object, G_Clit_O AND G_Valency_complnumb_O are used, e.g. in *Ho dimenticato di scrivere [scriverlo] prima. If a reference problem is caused by a missing/wrong clitic (e.g. "vederlo" instead of "vederti"), a C_coh_ref tag is added.</p>	a), b), d), token with the wrong, added or wrongly placed element c) verb	<p>a) puoi [chiamarla] {puoi chiamarmi} ; vorrei [mandarla] il mio curriculum {vorrei mandarLe il...}</p> <p>b) ho dimenticato di [scrivere] prima {ho dimenticato di scriverlo prima}</p> <p>c) [sono] le lezioni {ci sono le lezioni}; non problema {non c'è problema}</p> <p>d) il lago non [c'era] vicino all'appartamento {il lago non era vicino...}</p>	T
part of speech		G_POS	<p>This tag is supposed to identify the erroneous use of parts of speech (word classes). If a learner uses a noun instead of a verb, or an adjective instead of an adverb, for example, this tag is used.</p>	one token (wrong POS)	<p>ITA: *Tra cui (queste) l'inglese è la più facile; *Ho dimenticato di (come) parlare tedesco; *Ogni (I) bambini dovrebbero studiare; *Una signora molta (molto) simpatica qui (che) si chiama Maria; *Non ha sentito buono (bene); *Quando (Quanto) costa il corso?</p> <p>GER: *Ich freue mich für unsere [besucht] {Besuch}; *Ich bin sehr flexibel und [Mobilität] {mobil}; *Kannst du mich [Hilfe] {helfen}; *[Zusammenfassung] lässt sich sagen {zusammenfassend}</p> <p>CZE: *bylo to pro mě [potěšený] {potěšení}, *musíš myslet [rozumný] {rozumně}, zdravý život bez [nemocni] {nemoci}, doufám, že počasí bude [dobře] {dobře}</p>	R

MERLIN ANNOTATION SCHEME (II)
ORTHOGRAPHY TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
grapheme	general grapheme error	O_Graph_graphgen_O O_Graph_graphgen_Ad O_Graph_graphgen_Ch	In MERLIN, if more than one letter in one grapheme is wrong, it counts as one error. If there is more than one grapheme error in one token, all errors are annotated with the appropriate tags. This tag includes all errors concerning the choice of the correct grapheme, and it is used if: a) one or more graphemes is/are missing at the beginning or the middle of the word (but not in inflectional suffixes). This includes the simplification of consonant groups (O_Graph_graphgen_O) b) one or more graphemes is/are erroneously added. This includes the doubling of consonants and vowels (but not in inflectional suffixes) (O_Graph_graphgen_Ad) c) a grapheme is wrongly used instead of another grapheme (again, this does not apply to inflectional suffixes). Capitalization errors are not included here (--> O_Capit) (O_Graph_graphgen_Ch) LANGUAGE-SPECIFIC REMARK GER: <ß> and <ss>: If the learner uses the grapheme <ss> consistently in the whole text, no tags are applied (even in cases like: "Strasse", "Mass" etc.). In case of inconsistent use of <ss> and <ß> (e.g., "Straße" and "Strasse" in one text), the Swiss spelling counts as an error. GER/CZE: if the concerned vowel was part of the inflectional paradigm, errors in the stem are not tagged with O_Graph_graphgen RELATED TAGS This tag does not apply to graphemes with diacritics (CZE: á, é, í, ó, ú, ů, š, č, ě, ř, ž, ITA: è, ò, ì...); here, O_graph_act was used.	1 token	a) *těšila {těšila}, *vidme {vidíme}, *Pardubce {Pardubice}, *cera {dcera}, *děcký {dětský}; *wen {wenn}, *libe {liebe}, *sviluppo {sviluppo}, *leto {letto}, *viaggare {viaggiare}, *nove {nuove} b) *Peter {Petr}, *čtvrtek {čtvrtek}, *purst {prst}, *červenec {červenec}, *banek {bank}; *schpazieren, *eigentlich {eigentlich}, *mann {man}, *Monart {Monat}; *reggionale {regionale}, *occasione {occasione} c) *svadba {svatba}; *těším se na [ji] {ni}; *schreibt {schreibt}, *volte {wolte}, *Auslander {Ausländer}; *soziologia {sociologia}, esperienza {esperienza}, *mo {ma} Examples for grammatical errors (not annotated as orthographical errors): *píše tužku {tužkou}, *nese {neseš}; *has {hast}; *konnst {kannst}; *hilfen {helfen}	R/U
grapheme	transposition	O_Graph_trans	This tag is used for the transposition (interchanging) of two or more contiguous or non-contiguous graphemes (vowels and consonants). For annotation of proper names, see O_Graph_graphgen .	1 token	*kraští {kratší}, *asterix {asterisk}, *calvary {cavalry}, *cerel {celer}, *revelant {relevant}; *schriebst {schreibst}, *diene {deine}, *saulti {saluti}	R
grapheme	accent	O_Graph_act_O O_Graph_act_Ad O_Graph_act_Ch	This tag applies to all errors with diacritically marked graphemes (i.e., graphemes with accents). It is used in the following cases: a) the grapheme does not contain an obligatory diacritic (O_Graph_act_O) b) a diacritic is wrongly added (O_Graph_act_Ad) c) a diacritic is wrongly used instead of another diacritic (O_Graph_act_Ch) In all three scenarios, grammatical endings are not tagged with O_Graph_act. LANGUAGE-SPECIFIC REMARK ITA: In the case of "un pò" instead of "un po' ", it was decided to tag both O_Graph_act_Ad and O_Apostr_O RELATED TAGS O_Graph_act does not apply for errors involving the apostrophe (O_Apostr). For annotation of proper names, see O_Graph_graphgen .	1 token	a) *[kratši] {kratší}, *[dɔvod] {důvod}; *[e] andata {è}, Siamo [li] {li}, b) *Váčlav, ...*[Hà] chiamato; *[papà] {papa}; *un pò {un po'} c) úplný {úplný}; *perchè {perché}	R/U
capitalization		O_Capit	This tag is used for annotating: a) the use of lower case instead of upper case letters (and vice versa) in proper names (street names, names of institutions, states etc.) and (for GER) in nouns b) the use of upper case instead of lower case letters in all other categories of words (exception: beginning of sentences) c) the use of lower case instead of upper case letters in the beginning of sentences and the use of upper case instead of lower case letters after a comma d) the use of upper case letters in the middle of words (abbreviations do not count here) LANGUAGE-SPECIFIC REMARKS GER: "Du"/"Dir" (written in capital letters) in informal letters is not tagged as capitalization error. GER/ITA: Proper names are not tagged with O_Capit except for: names of countries, continents, oceans and names of German or Italian regions / federal states (endonyms). CZE: Proper names are not tagged with O_Capit except for: names of countries, continents, oceans, names of Czech regions and towns (endonyms) and Czech names of foreign towns and regions (exonyms), eg. Londýn, Drážďany	1 token	a) *evropě {Evropě}, *štědrý {Štědrý}, *Vietnamský {vietnamský}, *pardubice {Pardubice}; *via mazzini {Mazzini}; *in Maggio {maggio} b) *Staré {staré}, *Rodině {rodině}; *in Maggio {maggio}; *il testo per l'Italiano {italiano} c) *[byli] jsme ve Francii.; * [sie] waren in Frankreich, [Und] danach in Deutschland.; *solo, [Ora] non so il perchè; * era sereno. [poi] è nevicato. d) *[FahrkarTe]; *ich [wollTe] ...	U

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
word boundary		O_Wordbd_Split O_Wordbd_Merge	<p>This tag applies to:</p> <p>a) wrongly separated prefixes (O_Wordbd_Split) b) divided compounds (O_Wordbd_Split) c) other splitting (O_Wordbd_Split) d) wrongly joined prefixes (O_Wordbd_Merge) e) other wrongly joined words (e.g. articles) (O_Wordbd_Merge)</p> <p>RELATED TAGS If a compound is split and in addition contains a capitalization error, both the capitalization (O_Capit) and the word boundary error are annotated (e.g. Bewerber gespräch). If both the beginning of the first and of the second split word contain capitalization errors (e.g. *bewerber gespräch), only one capitalization error is to be annotated. If wrongly joined words involve a missing apostrophe (i.e. *daccordo instead of "d'accordo"), O_Apostr_O is also used. In MERLIN, O_Wordbd_Merge is used as general, overarching tag for all merges whereas G_Prep_Merge and G_Art_Merge are special cases of word boundary errors. That means that every G_Prep_Merge/G_Art_Merge is at the same time an O_Wordbd_Merge. For conflicts resulting from cases in which articles and prepositions are merged, G_Art_Merge is used. G_Art_Merge and G_Prep_Merge apply for merges of prepositions and articles with words from all word classes.</p>	1 or more token (depending on divided parts)	<p>a) před připraven; *a morale {amorale} b) *česko slovenská spolupráce {československá}, *nad zemní parkoviště {nadzemní}, *tele vize {televize}; *Schlafe zimmer, *jeden falls, *12 jährige, *das selbe; *da vero c) ne čekala {nečekala}, *[qui ndi] d) *dolesa {do lesa}, e) *takdobrý {tak dobrý}, *atak {a tak}; *Desweiteren, *Sogut; *daccordo</p>	R/U/T
abbreviation		O_Abbrev	<p>All orthographical errors in the use of abbreviations are annotated with this tag</p> <p>RELATED TAGS For the choice of a wrong abbreviation, V_semdenot is chosen For decisions on proper names, see O_Graph_graphgen definition</p>	1 token	<p>*s.m.s {SMS} *at. {atd.} ITA: *etc. {ecc.}</p>	
punctuation		O_Punct_O O_Punct_Ad O_Punct_Ch O_Punct_Pos	<p>This tag is applied for the punctuation marks: comma, full stop, exclamation mark, semicolon, quotation marks, hyphen, ellipsis, and includes the following cases:</p> <p>a) missing punctuation mark (O_Punct_O) b) redundant punctuation mark (O_Punct_Ad) c) wrongly chosen punctuation mark (O_Punct_Ch) d) wrongly positioned punctuation mark (O_Punct_Pos)</p> <p>RELATED TAGS This tag does not count for apostrophes, accents and diacritics (see O_Apostr, O_Graph_act)</p> <p>LANGUAGE-SPECIFIC REMARK Punctuation and addressing (GER): In some test rubrics (letters), a comma is missing after the greeting (B1 tasks). Therefore, a missing comma (after the greeting, before the name) was not annotated as a punctuation error.</p>	<p>a) whole sentence or clause up to the missing mark (exception: quotation marks: tag span is the citation) b), c) redundant/wrong punctuation mark d) wrongly positioned punctuation mark</p>	<p>a) *{českoanglický} slovník {česko-anglický}, [*Er kam nicht] aber er hat sich nicht entschuldigt. b) Ráda bych pokoj, s televizí. *La scuola che ho frequentato per due anni a New York e che ora non esiste più [,] sorgeva in un palazzo molto bello. {La scuola che ho frequentato per due anni a New York e che ora non esiste più sorgeva in un palazzo molto bello} *Rom, Paris[,] und Berlin gefallen mir sehr. {Rom, Paris und Berlin gefallen mir sehr.} c) Tak chtěla bych zeptat, jestli máte volný pokoj? *Penso le persone in Italia sono gentili [.] Ma purtroppo resto solo per cinque giorni [.] E poi devo ritornare a casa mia {Penso le persone in Italia sono gentili, ma purtroppo resto solo per cinque giorni e poi devo ritornare a casa mia.} *Komm her[?] Komm her {!} d) *Nevím, ale co je nového. *Ich weiß [,] nicht was es Neues gibt. {Ich weiß nicht, was es Neues gibt.}</p>	
apostrophe		O_Apostr_O O_Apostr_Ad O_Apostr_Pos	<p>This tag was used in the following cases:</p> <p>a) missing apostrophe (O_Apostr_O). If necessary, O_Wordbd_Merge was also used b) redundant apostrophe (O_Apostr_Ad) c) wrongly positioned apostrophe (O_Apostr_Pos)</p>	1 token, if missing: the token where the apostrophe should go (including token that is to be merged)	<p>a) * anchio {anch'io}, *un [po]/[pò] {un po'}, *daccordo {d'accordo}; [d{' }accordo] b) *Das ist Katrin's Grillstübchen. {Das ist Katrins Grillstübchen}; *Sono un'utente esperto. {un utente}</p>	

MERLIN ANNOTATION SCHEME (III)

INTELLIGIBILITY TAGS (EA2 - annotations available for core corpus)

SPECIFICATION	TAG DENOMINATION	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U= users; C=CEFR; T=text analysis)
intelligibility of text	G_Intelltxt_0 G_Intelltxt_1 G_Intelltxt_2	The intelligibility of each text is assessed. It is a general assessment referring to the whole learner production including the address, date, greeting, main text, farewell and possible other parts. The tag has the following specifications: a) The text is completely intelligible. Errors and unclear meanings might occur, but all sentences are comprehensible. G_Intelltxt_2 b) In the text, deviations occur (e.g. one or two sentences or parts of sentences are incomprehensible), but the context makes the meaning clear. The text as a whole is still intelligible. G_Intelltxt_1 c) The text is not comprehensible. More than one half of all sentences is not intelligible (e.g. because they are interrupted). G_Intelltxt_0	text	n.a.	R/C
intelligibility of sentence	G_Intells_0 G_Intells_1	This tag is used for sentences that can not be understood without problems. If the sentence is absolutely comprehensible (regardless of grammatical and other errors), this tag is NOT used. The tag has the following specifications: a) The sentence manifests deviations, but it is still interpretable. G_Intells_1 b) The sentence is completely incomprehensible (e.g. because it is not completed, there are a lot of lexical and/or grammatical errors). G_Intells_0 Addresses, dates, and other parts do not get the G_Intells tag, while the text body including the greeting, the main text, and the farewell do. RELATED TAGS The tag V_lexgrammarerror_incompr is similar, but it does not necessarily include a complete sentence. The G_Intells tag, however, refers to a whole sentence. If necessary,	sentence	a) GER: *[Diese Wohnung sind Bitte 2,3 km von Centrum, warum ich immer fahre mit meine fahrrad, und meine Beruf ist Kürche Hilfe, immer morgens, und Abend Arbeit.] CZE: *[Určitě, dobře si využijeme ten večer.] b) GER: *[meine wollten ist gut ich Besoche Kerche und centrem.] CZE: *[Děkuje za E-mail, že jsi pozval k narozeniním.] ITA: *Nell'ultimo periodo ho imparato tutti i giorni, perchè scriverò molte l'esame nel mesi seguenti e poi vuole fare la maturità anche., *[A queste cita di posto?]	R/C

MERLIN ANNOTATION SCHEME (IV)						
VOCABULARY TAGS (EA2 - annotations available for core corpus)						
Specification (1)	Specification (2)	TAG DENOMINATION	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
Formulaic sequence (FS)	collocation	V_FS_colloc	<p>DESCRIPTIVE 'EXISTENCE TAG'</p> <p>A collocation is a type of formulaic sequence (FS; definition: glossary). The tag is used if a sequence of words is recognised as a collocation independently of whether it is used in a semantically or formally correct or appropriate way. V_FS_colloc is thus used for all collocations in the corpus (correct & incorrect).</p> <p>REFERENCES GER: Duden Universalwörterbuch or: Duden online, Duden Stilwörterbuch; DWDS Kernkorpus des 20./21. Jh. --> min. 5 occurrences (cf. Nesselhauf 2005: 54) ITA: Treccani, http://www.treccani.it/enciclopedia/ricerca/ CZ: Příruční slovník jazyka českého (available also here: http://bara.ujc.cas.cz/psjc/search.php), Slovník spisovného jazyka českého (available also here: http://ssjc.ujc.cas.cz/), Slovník spisovné češtiny pro školu a veřejnost</p> <p>SOME DETAILS Typical constructions are: verb-object, adverb-adjective, adjective-noun. Also included in V_FS_colloc: - combinations involving content words + function words, such as: in der Sonne sitzen (but not: an/unter der Sonne sitzen, Burger 2007, Feilke 1996) or noun-preposition-noun combinations [in ITA: V_FS_coll_compeq] - sequences with delexical verbs (Funktionsverbgefüge); - comparative sequences, pair formulae and kinegrams [in contrast e.g. to Burger 2007]</p> <p>RELATED TAGS If there is an error in the use of a collocation tagged as V_FS_colloc, one or more MERLIN error tags are applied in addition to V_FS_colloc. Salutations/complimentary closes are tagged with the S_txt_grfw tag.</p>	FS (if discontinuous, only FS elements tagged)	<p>CZE: *dávát pozor, *mluvit pravdu delexical verbs: *předat někomu slovo, *navázat na něčí úspěch comparative sequences: *být chytrý jako liška, *chovat se jako doma pair formulae: *chtě něčtě, *horem dolem kinegrams: *krčít rameny, *mávnout nad něčím rukou, *pokývat hlavou, *pozvednout obočí</p> <p>GER: *Zähne putzen, *Klavier spielen, *eingefleischer Jungeselle, *hohe Stirn, *heute Morgen, *in der Sonne liegen, *Informatik mit Schwerpunkt XY Funktionsverbgefüge/delexical verbs: *eine Frage stellen, *zur Anzeige bringen comparative sequences *flink wie ein Wiesel, *stumm wie ein Fisch pair formulae: *klipp und klar kinegrams: *die Nase rümpfen, *die Achseln zucken</p> <p>ITA: *andare in vacanza, *mangiare un boccone in fretta, *giocare a palla Funktionsverbgefüge: *porre una domanda, *prendere una decisione, *mettere a disposizione, *prendere in considerazione comparative sequences: *muto come una tomba* pair formulae *chiaro e tondo* kinegrams *scrollare le spalle</p> <p>MERLIN examples GER: *... dass meine Tochter im April ein gesundes [Kind zur Welt bringt]; *Wie kann ich die [Freizeiten verbringen]; ... * Kann man sich leisten, nicht [in die Tiefe zu gehen]; ...; [Erfahrung im Umgang mit] Kindern und der Haushaltsführung; *[ich habe] vielen Stunden [Klavier gespielt]; *[den Teufel an die große grüne Wand malen]</p> <p>ITA: *[ho suonato il pianoforte] - *[ho suonato] per tante tante ore [il pianoforte],</p> <p>CZE: *[dávěj na sebe pozor], *[měj se hezky], *[hodně štěstí a zdraví], *[nabyla jsem dojmu], *[složil zkoušku], *[vyrazil ven], *[překonal jsem strach], *tam [se cítím jako doma], *[naplnit své potřeby]</p>	R/C/I
Formulaic sequence (FS)	collocation: ITA compound equivalent	V_FS_colloc_compeq	<p>DESCRIPTIVE 'EXISTENCE TAG'</p> <p>A compound equivalent is a type of collocation (see glossary). The tag is used if a sequence of words is recognised as a compound equivalent independently of whether it is used in a semantically or formally correct or appropriate way.</p> <p>RELATED TAGS If there is an error in the use of a compound equivalent tagged as V_FS_colloc_compeq, one or more MERLIN error tags are applied additionally.</p>	FS (if discontinuous, only FS elements tagged)	<p>[occhiali da sole], [ferro da stiro] [pesce spada], [vagone letto], [parola chiave], [pentola a pressione], [lente a contatto], [sala da pranzo], *[lista di desideri] {lista di nozze}</p>	R
Formulaic sequence (FS)	idiom	V_FS_idiom	<p>DESCRIPTIVE 'EXISTENCE TAG'</p> <p>This tag is used if a formulaic sequence (FS) is identified as an idiom (definition: glossary), independently of whether it is used in a semantically or formally correct or appropriate way.</p> <p>RELATED TAGS Idioms are sometimes not easily distinguishable from collocations (V_FS_colloc). Reference in case of doubt: monolingual dictionary, check figurative meaning (for German: Duden Universalwörterbuch or: Duden online; ITA: Treccani, CZE Příruční slovník jazyka českého (available also here: http://bara.ujc.cas.cz/psjc/search.php), Slovník spisovného jazyka českého (available also here: http://ssjc.ujc.cas.cz/), Slovník spisovné češtiny pro školu a veřejnost; If there is an error in the use of a compound equivalent tagged as V_FS_idiom, one or more MERLIN error tags are applied additionally.</p>	FS (if discontinuous, only FS elements tagged)	<p>CZE: *najít klíč ke štěstí, *mit černé svědomí GER: *etwas auf die lange Bank schieben, *blinder Passagier, *Morgenstund hat Gold im Mund. (proverb) ITA: *non cavare un ragno dal buco</p>	R
Formulaic sequence (FS)	communicative phraseologism	V_FS_commphras	<p>DESCRIPTIVE 'EXISTENCE TAG'</p> <p>This tag is used if a formulaic sequence (FS) is identified as a communicative phraseologism (CP), independently of its correctness or appropriateness. Includes common places, dicta (may include commonly used slogans from publicity), but no named entities (e.g. "das Weiße Haus")</p> <p>RELATED TAGS If a CP is at the same time a metacommunicative device, both C_FS_commphras and C_Coh_txtstruc are used Salutations, complimentary closes, addressing are NOT considered for V_FS_commphras, but for S_Txt_grfw.</p>	FS (if discontinuous, only FS elements tagged)	<p>GER: *Wie geht's, wie steht's? *Mach dir nichts draus. *Mir fehlen die Worte *... sag ich mal ...; *ich meine ...; *meines Erachtens ...; *ehrllich gesagt ...; *soweit ich weiß ...; *Rück endlich mit der Sprache raus, ... Abarakadabara Simalabim ... common places: *Was man hat, hat man. dicta: *Sein oder Nichtsein, das ist hier die Frage; *Nicht immer, aber immer öfter.</p> <p>Examples from MERLIN texts: GER: *... [und so weiter]; *...[meiner Meinung nach]...; *[Danke im Voraus] ...; *[Herzliche Glück wunsch] ...; *[Zum Glück] habe ich...</p> <p>ITA: *non so che dirti ..., *ascolta ...; *come dico sempre ...; *secondo me ...; *a mio avviso ... common places: *Quel che è fatto è fatto.</p> <p>CZE: *pokud vím, tak...; *mám na mysli..., *dle mého názoru..., *upřímně řečeno...; *věnujte pozornost...; *jak bylo řečeno výše...; *přejděme k dalšímu... common places: *Co se stalo, stalo se. dicta: *Méně je někdy více, *Vše má své výhody a nevýhody.</p>	R

Specification (1)	Specification (2)	TAG DENOMINATION	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
General	lexical & grammatical error(s) leading to incomprehensible expressions	V_Sequence_Lexgrammerr_incompr	This tag captures accumulations of meaning-form lexicogrammar errors within a sequence of words (also beyond a sentence end) that make the sequence incomprehensible as a whole. The tag is applied if it is impossible to specify and distinguish single errors reliably. RELATED TAGS V_Sequence_Lexgrammerr_incompr is NOT used for FS, but for free combinations of words. For FS, FS-specific tags are chosen (for an accumulation of form errors in FS: V_form_FS_incompr; for meaning errors: V_sendenot_fs; even both can be used if necessary). G_Intells: The focus of V_Sequence_Lexgrammerr_incompr is on the error dimension (in contrast to the intelligibility dimension). If necessary, both tags are applied.	whole unit	CZE: *Uřčitě, že pro kládný výsledek z něžého máme hodně pracovat a snažit se dosáhnout toho všem sebou. GER: *lohnt sich sicher lich, sich danach, zu [strenden], das ... (from Merlin corpus), *Ich wünsche mich meine Sohn alles guten zum Geburtstag (ex from Merlin texts, incomprehensible in context), *O.K. (ex from Merlin texts: single word as sentence, incomprehensible in context) ITA: *credo veramente che tutti i richieste siano ufuai che io posso darLa	R
General	non-existing form	V_form_word_fs_nonexist	This tag is used to annotate non existing words or FS. It refers to a) words that don't exist, i.e. non-words, invented word forms - in these cases it is not possible to identify any target language word that could be meant. b) FS that only appear to be FS, that are allegedly used as a chunk but do not really exist in the target language (possibly translations from L1/another language) c) use of words/FS that do not exist in the specific language and that are usually not used in it, i.e., words from other languages than the target language REFERENCES: GER: German DUDEN Universalwörterbuch or DUDEN online ITA: http://www.treccani.it/vocabolario CZE: Příruční slovník jazyka českého (available also here: http://bara.ujc.cas.cz/psjc/search.php), Slovník spisovného jazyka českého (available also here: http://ssjc.ujc.cas.cz/), Slovník spisovné češtiny pro školu a veřejnost. RELATED TAGS Tag is only used if it no other, more exact tag is suitable - cases like (GER) "Bewerberggespräch" instead of "Bewerbungsgespräch", "nebenagusten" instead of "Nebenkosten", ITA: "bracciare" instead of "abbracciare" or "coraggiare" instead of "incoraggiare" are not annotated with V_form_word_fs_nonexist, but with V_word_form_deriv/comp. Common foreign words the use of which is natural in the target language and which are listed as lemmas in the reference tools named above are not tagged.	word or FS	CZE: a) výsledky {průžek} {?} b) [trvalí] čas {trávili} c) urobit, i touch, freelance GER: a) *Kaus b) *wer will schon Staub essen (bite the dust) ... c) *... ist ein Menefrehgista ...; ... *hope (hoffen) ITA a) * passegere {meaning passeggiare} b): bisogna mangiare una mela acida c) : compra mlk e tomates	R
General	form errors leading to incomprehensible formulaic sequence	V_form_fs_incompr	This tag is used for formulaic sequences only. It is applied if in an FS there are so many form errors that it is rather difficult or impossible to reconstruct if and what is omitted, substituted etc. The FS is semantically correct/acceptable in the specific context although due to the accumulation of form errors the comprehensibility may be hampered or difficult.	FS (if discontinuous, only FS elements tagged)	GER: *[viel Öl zu heiß geben] {Öl ins Feuer gießen}; [ein Vögel zu Griechenland bringen] {Eulen nach Athen tragen} ITA: *ho pagato troppi occhie dalle testi --> ... un occhio della testa CZE: *[Vždycky dvejte na člověka není z hledisko v čem je.]	
Semantic error	denotation	V_sendenot_word_fs_0 V_sendenot_word_fs_1	Semantic errors' are instances in which a word/formulaic sequence is used semantically wrong so that the comprehensibility of the proposition is affected. The tag is applied to both single words and formulaic sequences. If it is applied to formulaic sequences, only errors regarding Idiomatic meaning are annotated (e.g. German: jemandem einen Korb geben). A further specification is made in the use of this tag: It regards the severeness of the semantic error. We distinguish between... 1 = deviation from hypothesized intended meaning is so minor that reader has no trouble substituting the correct word: V_sendenot_word/fs_1 0 = wrong, incomprehensible, hardly or not inferable from context: V_sendenot_word/fs_0 RELATED TAG If, for example, a part of the word/FS is changed, omitted etc, this is not a semantic, but a form error.	word or FS	(numbers in brackets behind sentences refer to error severeness) Word: CZE: *[využít si] života {užít si} života(1) GER: kauen {essen}, *sich die hand mit warmem Wasser verbrennen, *ich [mag] {möchte} neue Länder kennen lernen (1), ich konnte meinen Hund nicht [mitbringen] {mitnehmen} (1), *Ihr Baby [gewohnt]! {geboren???} (0), ...kannst du mir ein paar blaue Augen kaufen und [zurück bringen] {mitbringen} (1) ITA: venire instead of andare, imparare instead of studiare, abitare instead of stare/alloggiare (1) Formulaic sequences: CZE: *[zaměstnání na celou dobu] {zaměstnání na plný úvazek} (1) GER: *Das ist eine schwierige Zeit. Jetzt müssen wir alle [ins Gras beißen] (instead of: die Zähne zusammenbeißen) (0) ITA: *Gli occhiali sono costati un occhio della testa; per cui ne ho presi tre paia (0)	R
Semantic error	connotation (attitude)	V_semcon_att_word_fs	This tag is used if the word/FS does not have an appropriate (connotative) meaning with regard to the speaker's/writer's attitude in the given context; it reflects the emotional attitude of the speaker/writer towards the denotation. Such attitudes may be of the following kind, for example: - negative, pejorative, derogatory - ironic - euphemistic - jocular - others In order to verify the existence of such a connotative meaning of the word/FS, the context of use is taken into consideration. Reference tools: for German: DUDEN Deutsches Universalwörterbuch, for Italian: Treccani. For Czech: Příruční slovník jazyka českého (available also here: http://bara.ujc.cas.cz/psjc/search.php), Slovník spisovného jazyka českého (available also here: http://ssjc.ujc.cas.cz/), Slovník spisovné češtiny pro školu a veřejnost.	word or FS	Context-dependent tag (in other context, the following examples might be appropriate) Word: CZE: *[barák] {dům} GER: *[Köter] {Hund}, *[Alter] {Vater} ITA: *bagnarola {(vecchia) automobile/imbarcazione} FS: CZE: odejít navěky {zemřít} GER: *ins Gras beißen {sterben} ITA: *i miei vecchi {i miei genitori}	R

Specification (1)	Specification (2)	TAG DENOMINATION	DESCRIPTION	TAG SPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
Semantic error	precision	V_semimprec	<p>This tag has to be used for different cases that do not constitute real incorrect forms. Rather, they refer to lexical choices that lead to propositions that are in a way semantically imprecise, unusual, not native-like:</p> <p>a) The tag is used if a word/sequence of words is semantically acceptable and comprehensible but unusual, not precise, not specific, or with other words: if a unusual word/sequence of words is used to refer to something/someone. The tag refers mainly to the (hierarchical) paradigmatic semantic relations of the lexicon, as it may refer to the use of a hyperonym/generic term instead of a hyponym/specific term/sequence of words, or a holonym instead of meronym or viceversa or similar cases.</p> <p>b) The tag is used if a sequence of (adjacent) words is used although a specific, more precise term/sequence exists to express the same meaning; the formulation is semantically acceptable and comprehensible but imprecise, unusual, complicated (umständlich). This kind of error can also be described as "verbosity" (James 1998) or the use of paraphrases/circumlocutions. It refers to an error category that can be described as stylistic meaning infelicity (Hemchua/Schmitt 2006).</p> <p>Important: Use this tag only if you can exclude that the error is a semantic denotation or connotation error of a single word or a FS or even a lexical form error.</p>	<p>the word or whole sequence ALSO SEE "READ_explanatio ns_VOCAB" sheet!</p>	<p>a) GER: *[Ding] (instead of mentioning a concrete object); *eine [Liste] {eine Liste mit Wohnungsadressen/Kontakte zu Immobilienmaklern}; (example MERLIN texts) ITA: *vi devo chiedere qualche cosa {informazione}, *C'è qualcosa della Svizzera {di svizzero} che posso portarvi? CZE: doma má roztomilé {zvíře} {psa}</p> <p>b) GER: *[eine Firma, die Bücher macht] {Verlag}; ... im [Zug in der Nacht] {Nachtzug} ITA: *[per me ci sono tante cose nuove] {per me ci sono tante novità}, *la [gente inglese] {gli inglesi} (example MERLIN texts), *[sono un amante della lingua italiana e del paese] {dell'Italia o del Bel Paese} CZE: *[věci, které budou studované] (instead of studijní obory), vzdělání {dalo pro mě velkou pomoc} {pomohlo mi}</p>	R
Form error	word formation	V_Word_form_deriv V_Word_form_comp	<p>This tag is used if a single word has a wrong form, i.e. if the word form is deviating from the conventional form. Only errors regarding the lexical area of derivation and composition are considered:</p> <p>a) form errors regarding derivational aspects: wrong form related to the selection of pre-, in-, or suffixes. Inflection suffixes are not included here, thus no morphological errors regarding inflection are tagged with V_Wordform_deriv. Missing or wrongly selected affixes leading to unusual or inexistent words count as word formation errors, and so does the use of non existing affixes.</p> <p>b) form error regarding composition: errors in building compound words, e.g. wrong order of components, wrong or missing, linking elements (GER: 'Fugenelemente'), wrong POS selected for one or more parts of the compound: V_Wordform_comp</p> <p>For GERMAN: single letters are not tagged with this tag (exception: linking elements of compounds).</p> <p>RELATED TAGS Wrong affix leading to an erroneous meaning --> semantic error (V_semdenot_word, e.g. for German: *eindecken statt {zudecken}).</p>	<p>single lexical token</p>	<p>a) derivations: CZE: odpovídání {odpověď}, opravdivý {opravdivý} GER: *Suchung {Suche}, *[unheilsam] {unheilbar}, ITA: *bracciare instead of {abbracciare}</p> <p>b) compounds: CZE: životuschopný {životaschopný} GER: *Sprache Kurs {Sprachkurs}, *[italiensprachig] {italienischsprachig}, *[Türhaus] {Haustür}, *[Liebrie] {Liebesbrief} ITA: *ferro di stiro, *areoport</p>	R
Form error	formulaic sequence	V_FS_Form_O V_FS_Form_Ad V_FS_Form_Ch V_FS_Form_Pos	<p>This tag is only used if an FS is incorrect with regard to its formal elements.</p> <p>The following cases are distinguished: a) FS is incomplete, omission of one or more words: V_Form_FS_O b) addition of one or more words: V_Form_FS_Ad c) substitutions with regard to one or more of these aspects: c.1 wrong choice/substitution of one or more words; c.2 substitution of a part of an FS with parts of another FS = mixture of parts of more than one formulaic sequence (only for FS of the target language); c.3 substitution of morpho-grammatical features with others, e.g. use of plural form instead of singular, passive form instead of active, definite article instead of indefinite article: V_Form_FS_Ch d) wrong position of one or more words: V_Form_FS_Pos</p> <p>Regardless of the number of errors of each category within one FS, each of the four tags above can be used max. once per FS.</p> <p>RELATED TAGS If there are grammatical errors within the FS (e.g. in the inflection of nouns or adjectives etc.), the appropriate grammar tag has to be added (see "GRAMMAR" section of the annotation scheme).</p>	<p>_o: preceding token _ad: added tokens _ch: token(s) that would have to be substituted _pos: sequence containing the wrongly placed token(s)</p>	<p>a) CZE: je to [jen] příklad z mnoha {je to jen jediný příklad z mnoha} GER: *etwas auf [die] Bank schieben {etwas auf die lange Bank schieben}; *... ist meiner [Meinung], nicht ein großes Problem {meiner Meinung nach} ITA: *[carini] {mei cari}</p> <p>b) CZE: brát něco {doslova} vážně {brát vážně} / {brát doslova} GER: *Öi ins [heiße] Feuer gießen; *den Teufel an die [große] Wand malen ITA: *prendere due piccioni [neri] con una fava instead of prendere due piccioni con una fava</p> <p>c) CZE: mít [určitou polohu] ve společnosti {mít určité postavení ve společnosti}, [známkové] obléčení {značkové obléčení} GER: *einen Termin [machen] {einen Termin vereinbaren/ausmachen}; *{aus der Reihe fallen} (from: aus der Reihe tanzen & aus dem Rahmen fallen); *auf [keine grüne] Zweige kommen {auf keinen grünen Zweig kommen}; *Viele Familien wählen also für eine [Wohnung in Miete] {Wohnung zur Miete}; ...in Betracht [nehmen] {in Betracht ziehen} ITA: la stagione d'estate instead of la stagione estiva (ex from Merlin texts), carini instead of miei cari (ex from Merlin texts - substitution of cari), non cavare un ragno dal buco instead of non cavare dei ragni dal buco, sono alla vostra disposizione instead of sono a vostra disposizione</p> <p>d) CZE: Kdo jinému [kopá jámu], sám do ní padá. {Kdo jinému jámu kopá, sám do ní padá.} GER: *Der Apfel fällt [vom Baum nicht weit.] ITA: *tanti [saluti cari] {tanti cari saluti}</p>	R

MERLIN ANNOTATION SCHEME (V)						
COHERENCE/COHESION TAGS (EA2 - annotations available for core corpus)						
Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
connectors	connector accuracy	C_Con_accur_O_0 C_Con_accur_Ad_0 C_Con_accur_Pos_0 C_Con_accur_Ch_0 C_Con_accur_O_1 C_Con_accur_Ad_1 C_Con_accur_Pos_1 C_Con_accur_Ch_1	This tag involves an error definition and a judgment about the severeness of the error by the annotator. ERROR DEFINITION (target language modification) Incorrect use of connectors can refer to ... a) the erroneous choice of a connector (C_Con_accur_Ch) b) the position of the connector (C_Con_accur_Pos) c) the omission of a connector where one has to be used (C_Con_accur_O) d) the addition of a connector where in L2 there is none (C_Con_accur_Ad) SEVERENESS OF ERROR Annotators considered if the connector use is: a) incorrect/inacceptable (C_Con_accur_Ch/Pos/O/Ad_0) b) not entirely correct, but acceptable/no misunderstanding caused, (C_Con_accur_Ch/Pos/O/Ad_1) CONNECTOR LISTS There are predefined lists of connectors for German, Czech, and Italian that annotators referred to. RELATED TAGS An error in the use of connectors is both a coherence AND a grammatical error. The grammatical side of this error phenomenon is tagged with G_conj. Thus, every connector error is annotated with two tags at least.	connector; in case of missing connector, whole clause including punctuation mark	GER: a) *dort gibt es viele Studenten [als] die Miete nicht sehr hoch ist b) * Ich will auch Istanbul besuchen, [weil] schicke mir bitte Informationen. c) *[Ich fände es am besten eine Möglichkeit gäbe,] eine Unterkunft in einer Gastfamilie zu bekommen. d) *[Für] was die Familien angeht, ... ITA: a) *Gli rivolgo allo scopo [che] ho qualche domanda. *Sarà meglio [quando] uno di voi mi chiama e parliamo di questa questione" *Il mio capo ha organizzato una riunione. Purtroppo un altro giorno per il nostro incontro? b) *La mia famiglia gioca volentieri a pallavolo, non [però] c'era nessun possibilità nella Residence" d) *Il mio titolo di studio è l'insegnante e per questo motivo mi piace lavorare [anche] con i bambini, organizzare le gite e l'altro divertimento CZE: a) Chtěla jsem se zeptat pokud máte parkoviště protože příjdu s autem. b) Přinese pokud nějaké jídlo, můžeme mít oběd. c) [Potřebovala bys pomoct,] klidně napiš. d) Pokud přijde-li, budu rád.	C,R,T
coherence	content jumps	C_Coh_jump	A content jump is meant to describe a point in the learner production where the reader gets the impression of a sudden turn of the topic of the text. A new topic is introduced by the author without the appropriate linguistic means of preparing such a change of content. This tag is directly derived from the CEFR coherence/cohesion scale ('jumpiness'). It is completely dependent on the perception of the annotator. The annotation of content jumps is part of the attempt at validating CEFR scales (>>research section on interface) Content jumps were not tagged if annotators had the impression that the content jump would be completely avoided if only the learner had inserted a paragraph. C_Coh_jump is used only if the passage in the text needs an extra explanation to be understood.	Annotate the first token in the sentence where the content jump begins	GER: 1) *In Istanbul wolle ich gefahren. Aber du weißt ja, unser Türkisch sie haben uns nicht richtig verstanden. [Maria ist ruhig geblieben.] [Tut mir leid! Ich war total nervös.] 2) *Ich komme morgen zu besuch. [Meine Familie geht es gut. Begrüßen dir von meine Familie.] [Ich mag nicht Tiere, weil ich keine Zeit habe.] [Ich gehe in den Deutschkurs] (...) 3) *Ich habe am Wochenende deine Brief bekommen. Das ist schön dass Anna in den Kindergarten und Max ist in der dritten Klasse. [Ich möchte eine Hasen haben, aber ich habe Allergie für die Haar.] [Ich mache meine Sprachkurs, aber die Prüfung ist nächste Woche, dann kann ich zu dir kommen"] 4) *Ich freue mich dass bei dir alles geht gut! [Ich habe etwas nachgedacht, was mag ich am liebsten.] 5) *Im Moment wir habe viel Stunde für arbeiten. [Wir wollte unbedingt nächste Monat Besuch bei dir, und zusammen ein Party machen, weil meine Sohn Geburtstag habe. Wann wir habe Zeit?] [Wie hoch dich und unsern Familien geht.] 6) *[Mit gutem Herzen empfehle ich allen, seine Bücher zu lesen. Man konnte schnell etwas von ihm vor allem im Internet finden dessen zahlreiche Gedichte schmücken die ungarische Kultur.] 7) *Ich bin verheiratet und habe ich 3 Kinder. [Wir arbeiten bis 04:00 Uhr.] ITA: 1) *Molte cordiali felicitazioni per il tuo ottimo voto. Durante gli ultimi tempi aiutavo a mia madre degente in una casa di cura per persone anziane. [...] Un cambiamento non è probabile. Forse c'è la possibilità per te di lavorare qualche mese alla pari all'estero per mantenere le tue conoscenze linguistiche [...]. Penserò in ogni caso a te. 2) *Io non so cosa voglio fare. Accanto il lavoro all'albergo mi interessano anche gli esami per diventare una professoressa. Tutto sommato, con il tuo esame superato c'è qualche cosa da fare. CZE: *Prázdňinové kurzy češtiny jsou zajímavé. Chci se ucházet o kurz a asi stipendium. Kolik to stojí? Kdy začíná kurz? [To je kouzelný, že zůstan chvilku v Praze.] Můžeš poslat mě toho inzerát?	C,R,T,U
coherence	reference	C_Coh_ref	Reference is understood as a property of items which "instead of being interpreted semantically in their own right, they make reference to something else for their interpretation." (Halliday/Hasan 1989: 31). In MERLIN, reference errors regard: - inconsistency between the intersentential and the intrasentential reference - disagreement between the presupposed items and referents. Reference errors give rise to reference ambiguity and inconsistency. The C_Coh_ref tag was used for reference errors which... a) involved unclear/wrong/missing reference of pronouns, names, adverbs, (in)definite articles b) occurred within (endings of) verbs, when there was inconsistency between the presupposed number of person/the implicit subject of a sentence an the inflected verb	part of text that causes reference problem, e.g. the wrong pronoun or noun/name	1) *wie geht es [seine] (=deine) Frau, was ist der Name [sein] (=dein) Babys" 2) *meiner Küssen für [ihre] (=deine) Kinder 3) *bringe ich ihm dieser Dienstag [es] zu dir 4) *Wir hoffe dich und [unseren] (=eurer/deiner) Familien geht. 5) *Man muss ja nicht alle Traditionen obligatorisch mitmachen, die [uns] nicht passen." (man-uns) 6) *Mit gutem Herzen empfehle ich allen, [seine] (kein Bezug!) Bücher zu lesen. Man konnte schnell etwas von [ihm] (kein Bezug!) vor allem im Internet finden [dessen] (kein Bezug!) zahlreiche Gedichte schmücken die ungarische Kultur. 7) *Die Frage ist sehr zusammengesetzt, [es] lässt sich nicht so einfach beantworten. 8) *aber du weißt ja, unser Türkisch sie haben uns nicht richtig verstanden. [Maria] ist ruhig geblieben." (kein Bezug von "Maria") 9) *Es ist wichtig nicht den eigenen Kultur zu verlieren, [nun] unsere Wurzeln. ITA: 1) *Spero che la vostra [=tua] famiglia anche è sana 2) *Se glielo non [te lo] pagano devi lavorare in una ditta nel tempo libero 3) *Ho 29 anni. (Ho dimenticato di scrivere [lo] prima), *Come stai? Che cosa fai? Ha anche [tu] molto lavoro? 4) *Che lavoro voglio fare?" for "che lavoro vuoi fare"; *Ti auguri un splendido periodo" for "ti auguro uno splendido periodo" CZE: 1) *pani, jehož se ptal {jíž}, *dal jsem to jeho bratrovi {jějimu} 2) Budeš mít narozeniny? Jaký dárek si přejete?	R, T

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
coherence	metacommunicative devices	C_Coh_txtstruct	<p>If the author of a text speaks about the text in the text, this is considered a device to structure the text (also: 'metacommunicative device'). These means are used to explicitly refer to anterior or posterior parts of the text. They serve to directly guide the recipient.</p> <p>RELATED TAGS In MERLIN, C_Coh_txtstruct is not used to identify errors. If a word or a sequence of word that structures a text/talks about a text is used in an incorrect way, other tags are used with the C_Coh_txtstruct tag. If is a connector, C_Con_accur is applied in addition to C_Coh_txtstruct. An example (ITA) is: the use of "prima" instead of "Prima di tutto" must be tagged as C_Coh_txtstruct and in addition as C_Con_accur_Ch_O). Also, the incorrect use of text-structuring means might cause reference problems (C_Coh_ref)</p>	the means that structures/talks about the text	<p>GER: "im Folgenden", "zusammenfassend", "abschließend", "wie bereits erwähnt/angedeutet", "erstens, zweitens, drittens", "wie wir besprochen haben/wie besprochen", "dazu will ich noch erwähnen, dass..."; "im Folgenden", "im folgenden Bericht/Abschnitt/Text/..."; "meiner Meinung nach", "wie ich schon erwähnt habe/wie schon erwähnt", "zusammenfassend", "meines Erachtens", "also alles insgesamt", "wie ich schon erklärt habe"; "Ich möchte dich noch etwas fragen", "Jetzt wechseln wir das Thema"; "ich frage mich, ob..." etc.</p> <p>ITA: insomma, in conclusione, in definitiva, anzitutto /innanzitutto/ prima di tutto/ per prima cosa; in primo luogo, in secondo luogo, infine, in altre parole, in breve, inoltre; si osservi poi; si noti, in particolare, che;...</p> <p>CZE: zároveň, přitom, nadto, navíc, pak, dále, potom, především, konečně, nakonec, nejprve, zaprvé, zadruhé, zatřetí,... (za prvé, za druhé, za třetí,...), za další, další, ani - ani, jednak - jednak, na jedné straně - na druhou stranu; eventuálně, jednoduše, krátce, naopak, popřípadě, prostě, přesněji, případně, respektive, tedy, totiž, tudíž, vlastně, zkrátka,...</p>	R, T

MERLIN ANNOTATION SCHEME (VI) SOCIOLINGUISTIC APPROPRIATENESS TAGS						
Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
text-type specific	greetings/farewell	S_Txt_grfw	<p>DESCRIPTIVE 'EXISTENCE TAG'</p> <p>All greetings at the beginning of letters/emails and all complimentary closes (farewell phrases or words) at the end of these texts are tagged with S_Txt_grfw regardless of correctness. If the learner text is not a letter/email, the tag is not applied.</p> <p>Greetings precede the text body of the letter, while complimentary closes (=farewell phrases or words) follow it at its very end.</p> <p>If there is more than 1 tag S_Txt_grfw, please include them all in 1 tag (e.g., [Un abbraccio, tanti saluti, Maria])</p> <p>RELATED TAGS</p> <p>This tag can be considered a concrete form of S_Txt_opcl, as greetings open and close letters/ emails. It is a tag used in most texts that are letters/emails.</p> <p>S_Form_gen is used in addition if a greeting is inappropriate, i.e. too formal/not formal enough (e.g., someone saying "Altes Haus!" or "Hallo Herr Präsident!" in a formal letter, or someone writing "cordiali saluti, Mario" in a letter to a friend).</p> <p>S_Form_addr is only used if there is a clear violation of the addressing rules and this is explicit in the greeting/farewell (like in "Hallo du, Frau Lehrerin").</p>	greeting expression and name (including punctuation mark)	<p>GER: "[Hallo Maria]", "[tschüß Herr Meier]"</p> <p>CZE: [Ahoj Davide]; [Dobry den Pane ředitel hotelu]</p> <p>ITA: *[Ciao Francesco] *[Tanti saluti, Maria]</p>	C
text-type specific	opening and closing formulae	S_Txt_opcl	<p>DESCRIPTIVE 'EXISTENCE TAG'</p> <p>Learners might insert opening and/or closing forms in their texts to introduce their letter/email or to make it clear that they are finishing it off. These language forms are marked with this tag regardless of errors. Of course the expressions might contain errors that require one or more of the error tags in this scheme.</p> <p>You will typically find openings and closings at the end of a letter/email, before the complimentary close, or at the beginning of such a text, after the salutation. The degree of formulaicity of these text structural means might not be very high.</p> <p>Tag is applied only in letters/e-mails, not in essays and other text genres.</p> <p>RELATED TAGS</p> <p>C_txtstruct: Opening and closing forms of letters are means to structure a text. Therefore, use C_Coh_txtstruct as well.</p> <p>Greetings are not part of S_Txt_opcl; they get the more specific S_Txt_grfw tag</p>	whole formula or sentence including punctuation mark	<p>GER: *Vielen Dank für Deinen Brief. Ich habe mich sehr gefreut. *Ich würde mich freuen, bald von dir zu hören. CZE: *S přáním hezkého dne, *Mejte se hezký, *Uvidím tě později ITA: *Nel frattempo ti invio cordiali saluti (letter to a friend); *Aspetto la sua risposta al più presto; *Come stai?; *Scrivimi presto</p>	C, U
formality	general: inadequate formality	S_Form_gen	<p>An important aspect of sociolinguistic competence is the choice of language forms the formality of which is adequate for a given task/in a given situation (diaphasic [register]/diastratic [style] variation dimensions, Koch/Oesterreicher 1990). Formality can be understood as a dimension of style/register. This means that S_Form_gen is more an appropriateness tag than a "real" error tag. If formality is concerned, S_Form_gen takes into consideration the context-dependent, connotative semantic aspect of the lexical dimension (e.g. saying "Bude" for "Zimmer" in a German formal letter).</p> <p>The S_Form_gen tag concerns language forms with an inadequate formality with respect to the requirements of the task. The differing formality levels of the MERLIN tasks require the choice of an adequate style. For example, in a formal letter, it is not acceptable to use vernacular language forms. All Merlin tasks have a defined category of formality (formal - informal which is accessible online via the tasks/ALTE-grid based task descriptions).</p> <p>Thus, formal language in informal tasks (and vice versa) is tagged with S_Form_gen. Reference tools (see other sections of annotation scheme) were used for crosschecking formal, informal or even substandard forms according to the diasystem of the 3 project languages, if possible. A second help is textual inconsistency - inductive text analyses revealed that often, inappropriate formality is recognizable because a single expression's formality 'is not in line with the rest of the text. In many cases, annotators' intuition was the decisive factor for this tag. In cases of doubt, annotators discussed single cases.</p> <p>S_Form_gen is also used for inappropriate greetings ("Hi Professor!").</p> <p>RELATED TAGS</p> <p>S_Form_gen is not used for formality problems resulting from errors in addressing people (Like "Lieber Professor, [gibst du] mir bitte einen baldigen Termin"): here, S_Form_Addr is used. S_Form_gen is more general than S_Form_Addr.</p>	the whole inadequate sentence or part of the sentence	<p>introducing a letter to a friend with ...</p> <p>GER:*[Sehr geehrter Marco], - [Willkommen in „Stadt X“] (task: Bericht über Wohnungsmarkt); *Ich suche eine [Bude] zum [Pennen] (hypothetisches Beispiel; Schreiben an potenziellen Vermieter)</p> <p>CZE:*[Ahoj pane řediteli]</p> <p>ITA:*[Egregio Andrea], *[ti ringrazio cordialmente della tua gentile lettera del 12 m.s.]</p>	T, R
formality	specific: inadequate addressing	S_Form_addr	<p>S_Form_addr is a specification of S_Form_gen.</p> <p>It also regards formality, but it exclusively refers to sociolinguistically inappropriate addressing forms. Italian, Czech, and German have linguistic specifications for informal and formal/polite addressing. If a learner does not adhere to the language-specific rules in completing MERLIN tasks, S_Form_addr is used.</p> <p>This tag has a very specific focus. S_Form_addr only refers to "tu/Lei", "Du/Sie" and "ty/Vy". For CZE/ITA: as ty/Vy can be included in the verb form, this tag refers also to the inappropriate use of these forms; it is applied for the use of imperative forms in all three languages where a personal pronoun does not appear. For GER: A polite plural with "ihr" instead of "Sie" is tagged as an error.</p> <p>The envisaged formality level of the tasks can be found in the online task descriptions.</p> <p>RELATED TAGS</p> <p>S_Form_gen is used for all other cases in which the formality of the learner text is not appropriate for the given task (e.g., substandard).</p>	2 tokens (=form of addressing pronoun, tu/Lei, du/Sie, ty/Vy + verb form) or 1 token (if pronoun included in the verb form)	<p>in a formal letter:</p> <p>ITA: "Mi [puoi] dare informazione sulle condizioni?"</p> <p>GER: "Kann ich Informationen von [euch] bekommen?"</p> <p>CZE: "Prosim, [máš] další informace pro mě?"</p> <p>in an informal letter:</p> <p>ITA: "[Vi] ringrazio per la tua email."</p> <p>GER: "Kommen [Sie] nächste Woche mich besuchen?"</p> <p>CZE: "Bylo by dobré, kdybyste přijela do Drážďan."</p>	T, R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
variation (single languages)	ITA: Lexicalised clitics	S_Var_clit	<p>DESCRIPTIVE 'EXISTENCE TAG', ITALIAN</p> <p>This tag is used for certain pronoun verbs in Italian. Pronoun verbs ("verbi pronominali") have the pronominal particle/clitic as constituting element. S_Var_clit is used for a special group of pronoun verbs that have 'ce, la, ne, le' as pronominal component ('lexicalised clitics', verbi procomplementari/pro-complement verbs (Russi 2008)). Lexicalised clitics are forms like: <i>farcela, non poterne più, dirla, andarsene, infischarsene, smetterla, finirla, ...</i></p> <p>These lexicalised clitics belong to oral variants. The phenomenon, though, can be observed more and more often also in written language. All lexicalised clitics regardless of their correctness are tagged. If the learner made a mistake in a construction with a lexicalised clitic, one or more MERLIN error tags are applied in addition to S_Var_clit.</p> <p>Clitics to be tagged are: -ci, -la, -le, -ne, -cela, -cene, -cisi, -sela, -sene</p> <p>Main reference: Russi, Cinzia (2008): Italian Clitics: An Empirical Study. Berlin usw.: Mouton/de Gruyter.</p>	clitic plus verb and other parts of the construction, whole formulaic sequence	<p>...-ci: arrivarci (arrivarci a capire), averci (ci ho paura), cascarci, correrci (ci corre una bella differenza)</p> <p>...-la: contarla, farla (la fa a tutti), farla franca, finirla, piantarla, spuntarla; smetterla</p> <p>...-le: darle, prenderle</p> <p>...-ne: darne; volerne e poterne, solo in locuzioni), farne (farne di tutti i colori), volerne (non volermene)</p> <p>...-cela: avercela, mettercela, farcela</p> <p>...-cene: volercene, correrene</p> <p>...-cisi: mettercisi</p> <p>...-sela: cavarsela, cercarsela, contarsela, darsela (darsela a gambe), farsela (con), filarsela, godersela, intendersela, passarsela (passarsela bene),</p> <p>...-sene: andarsene, fregarsene, intendersene, restarsene, starsene, tornarsene,</p> <p>POLIREMATICHE: darci dentro, dormirci sopra, mettercela tutta, venirme fuori, prenderci gusto, farsela sotto, vedersela brutta, venirsene fuori (con)</p>	R
variation (single languages)	ITA: duplicazione dei pronomi	S_Var_duppron	<p>ITALIAN</p> <p>In colloquial spoken Italian, the personal pronoun is sometimes used twice in a redundant way in order to express the personal involvement (raddoppiamento clitico/ duplicazione pronominale). In written language, this doubling is not appropriate because it is considered too colloquial.</p> <p>The following structures are tagged with S_Var_duppron:</p> <p>a) "a me mi xyz" (and similar, like "a noi ci"...)</p> <p>b) "ne xyz di", e.g. "ne racconta di storie"</p> <p>c) 'pronomi benefattivi' like 'mi bevo una birra'</p> <p>RELATED TAGS</p> <p>Dislocations can also be a form of pronoun redundancy ["Il libro te lo do domani"]. However, dislocations are also used in the written language, and they are acceptable language forms [existence tags]. So S_synstr is used instead of S_Var_duppron for all cases not mentioned here (a-c)</p> <p>G_Clit_Ad (grammar) is used as well for this phenomenon.</p> <p>As the use of this structure is not appropriate with regard to the formality level required by most or all MERLIN tasks, S_Form_gen is added.</p>	pronoun+ redundant pronoun + verb; objects excluded; tagspan might be discontinuous	<p>a) 1) [A me mi piace]...; 2) [A lui] non [gli lascio] nulla. 3) Tu [a me] non [mi puoi] mica dare del cretino come se nulla fosse! 4) Un tool che [a noi ci] ha molto colpito perché è stato progettato da Vincenzo di Nicola, un giovane imprenditore italiano, è GoPago</p> <p>b) "[Ne racconta di] storie!"</p> <p>c) "[Mi bevo] una birra", "[Mi vedo] un film"</p>	R
variation (single languages)	ITA: marked syntactic structures	S_Var_synstr	<p>DESCRIPTIVE 'EXISTENCE TAG', ITALIAN</p> <p>Italian has some particular syntactic structures that can be used to emphasize parts of the sentence. They do not count as errors and they are tagged independently of their correctness.</p> <p>Special syntactic structures that are annotated in MERLIN are:</p> <p>a) <i>Frase scisse</i> (left sentences): The 'frase scissa' is a proposition which is introduced by the verb "essere", followed by the constituent which is re-placed, and followed by a pseudo-relative clause. The rhematic element is moved to the left, and the pseudo-relative expresses the theme. Annotate the whole part of the sentence in which this phenomenon appears.</p> <p>b) <i>Dislocazioni a sinistra</i>: A constituent other than the subject is moved left. It must be taken up by a clitic pronoun, by a free pronoun or by a nominal syntagma.</p> <p>c) <i>Dislocazione a destra</i>: A constituent (thematic) is moved right, preceded by a clitic pronoun.</p> <p>d) <i>c'è presentativo</i> (special kind of cleft sentence): The element in the copulative phrase must be in the extralinguistic context.</p>	a) from the form of "essere" to the end of the sentence/part of sentence b) - d) from the beginning to the end of the sentence/part of sentence	<p>a) 1) sei tu che hai detto questo. 2) Sono le foto che mi fanno pensare alle vacanze dell'anno passato. 3) È lui che ha copiato, io non c'entro prof! 4) Sei tu che ti devi adattare al suo funzionamento, non il contrario.</p> <p>b) 1) Che non sarei venuto, lo sapevi benissimo. 2) La spesa l'ho fatta ieri quindi oggi sono libera. 3) Di casi simili ne ho visti parecchi, lasci fare a me.</p> <p>c) 1) Ne voglio parlare con te, dei miei problemi. 2) Non preoccuparti! Lo portiamo noi, il vino! 3) Ci vado domani in piscina, oggi non ho tempo. 4) Possibile che non te la ricordi, quella canzone che cantavamo da piccoli?</p> <p>d) 1) C'è mia cugina che ti vuole parlare. 2) C'è Andrea che ti cerca. Cosa gli dico? 3) C'è il bambino che sta male, devo chiamare in ufficio.</p>	R
variation (single languages)	ITA: che polivalente	S_Var_che	<p>ITALIAN</p> <p>The so-called 'che polivalente' means instances in which "che" as introduces subordinates which in standard Italian have more precise conjunctions. "Che polivalente" marks a colloquial style. In addition, the "che" polivalente is used with the indicative which often decides about its acceptability. The "che polivalente" is only marked if used incorrectly.</p> <p>The "che polivalente" introduces mainly</p> <p>1) frasi di significato esplicativo-consecutivo: <i>viene che ti pettino</i></p> <p>2) frasi causali: <i>vai a dormire che ne hai bisogno</i></p> <p>3) frasi consecutivo-presentative : <i>io sono una donna tranquilla che sto in casa, lavoro</i> (Sornicola 1981: 70-71)</p> <p>4) frasi relative temporali: <i>maledetto il giorno che ti ho incontrato</i></p> <p>5) frasi finali: <i>fai in modo che è tutto pronto al mio arrivo</i></p> <p>6) frasi in cui che ha valore enfaticamente-esclamativo: <i>che sogno che ho fatto</i> (Berruto 19984: 69)</p> <p>7), frasi pseudorelative: <i>li vedo che scendono</i> (Berruto 19984: 69)</p> <p>[categorization: http://www.treccani.it/enciclopedia/che-polivalente_%28Enciclopedia_dell%27Italiano%29/]</p>	whole sentence	<p>1) [Vieni qui che ti voglio dare qualcosa.] {in modo che/perché}</p> <p>2) [Ho sentito cose che non avevo fatto caso] {a cui non}</p> <p>3) [Il paese che sono stata] {in cui/dove}</p> <p>4) [Vai a dormire che ne hai bisogno]</p> <p>examples partly from Treccani: http://www.treccani.it/enciclopedia/che-polivalente_%28Enciclopedia_dell%27Italiano%29/</p>	R

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
variation (single languages)	GER: HS word order after weil	S_Var_wowell	<p>GERMAN</p> <p>In Merlin, if after the subjunctive "weil" there is a verb in the second position (Verbzweitsatz), this is tagged with S_Var_wowell. This phenomenon has been termed 'epistemic weil' (Keller 1998). More than a question of right and wrong, this phenomenon is regarded as an expression of diasystematic variation in MERLIN; it is regarded as a matter of acceptability more than of correctness. Therefore, S_Var_wowell belongs to the Sociolinguistic Appropriateness section and not to the Grammar section.</p> <p>RELATED TAGS</p> <p>G_Wo_wosubcl is additionally used on EA1 level for these phenomena.</p> <p>References:</p> <p>Pasch, Renate (1989): Adverbialsätze - Kommentarsätze - Adjungierte Sätze. Eine Hypothese zu den Typen der Bedeutungen von weil, da und denn. In: Motsch, Wolfgang (Hg.) (1989): Wortstruktur und Satzstruktur. Berlin: Akademie-Verlag. S. 141-158.</p> <p>Keller, Rudi (1993): Das epistemische weil. Beobachtungswandel einer Konjunktion. In: Heringer, Hans Jürgen/Stötzel, Georg (Hgg.) (1993): Sprachgeschichte und Sprachkritik. Festschrift für Peter von Polenz zum 65. Geburtstag. Berlin, New York: de Gruyter. S. 219-247.</p> <p>Konopka, Marek (2005): Topologie komplexer Sätze und Textverstehen. Zur Stellung von Verbletztsätzen mit weil. In: Blühdorn, Hardarik/Breindl, Eva/Waßner, Ulrich Hermann (Hgg.) (2005): Text - Verstehen. Grammatik und darüber hinaus. Berlin: de Gruyter. S. 108-124.</p> <p>Ravetto, Miriam/Blühdorn, Hardarik (2011): Die Kausalkonjunktionen denn, weil, da im Deutschen und perché, poiché, siccome im Italienischen. In: Ferraresi, Gisella (Hg.) (2011): Konnektoren im Deutschen und im Sprachvergleich. Beschreibung und grammatische Analyse, Tübingen: Narr. S. 207-250.</p> <p>Breindl, Eva (2012): Er ist sympathisch, weil menschlich - Weil als koordinierende Konjunktion? In: Konopka, Marek/Schneider, Roman (Hgg.) (2012): Grammatische Stolpersteine digital. Festschrift für Bruno Strecker zum 65. Geburtstag. Mannheim: Institut für Deutsche Sprache. S. 153-158.</p> <p>Mroczynski, Robert (2012): Grammatikalisierung und Pragmatikalisierung. Zur Herausbildung der Diskursmarker wobei, weil und ja im gesprochenen Deutsch. Tübingen: Narr. (Tübinger Beiträge zur Linguistik 530).</p> <p>Thielmann, Winfried (2014): Weil - zur Funktionalität eines 'kausalen' Ausdrucks. In: Deutsche Sprache. Zeitschrift für Theorie, Praxis und Dokumentation 42 1. S. 65-77. http://hypermedia.ids-mannheim.de/call/public/sysgram.ansicht?v_id=1607</p>	whole subordinate clause beginning with "weil" and ending with a punctuation mark (including punctuation mark, or comma, if another subclause follows)	1) Ich habe Hunger, [weil es ist ja auch schon ganz schön spät.]	R
variation (single languages)	GER: Abtönungspartikeln	S_Var_partik_0 S_Var_partik_1	<p>Combination of DESCRIPTIVE 'EXISTENCE TAG' and 'ERROR TAG'</p> <p>TAGS</p> <p>All listed particles are tagged ('existence tag') regardless of their correctness. Then, an additional tag extension specifies whether the use is target-language-like or not ('error tag').</p> <p>S_Var_part_0 = if the use of the particle is inappropriate or wrong or just unusual (e.g. Was willst du DOCH eigentlich? Was willst du ÜBERHAUPT eigentlich?).</p> <p>S_Var_partik_1 = if the use is appropriate/acceptable or perfectly fine.</p> <p>LIST OF PARTICLES</p> <p>halt ("Dann geh halt"); ja ("Das ist ja schön, das Buch"); eben ("Genau das meinte er eben nicht"); denn ("Was willst du denn bloß?"); doch ("Luise hat doch Bescheid gesagt, dass sie nicht kommen kann"); aber ("Das ist aber ein toller Hund"); ruhig, ("Geh ruhig schon vor"); nur ("Was hat er nur?"); mal ("Geh schon mal vor"); schon ("Das ist schon eine ganz schön komplexe Aufgabe"); überhaupt ("Was meint er denn überhaupt?"); vielleicht ("Das ist ja vielleicht ein Schmuttelwetter"); eigentlich ("Was meinstest du eigentlich mit der Bemerkung?"); wohl ("Carsten spinnt ja wohl").</p> <p>These particles must not be mistaken with homographic adverbs (cannot be asked for; refer to the whole sentence)! For example: "aber": "Ich würde ja gerne gehen, aber ich kann nicht"; or "vielleicht": "Vielleicht kommst du ja doch". --> no particles</p> <p>References:</p> <p>Waltereit, Richard (2006): Abtönung. Zur Pragmatik und historischen Semantik von Modalpartikeln und ihren funktionalen Äquivalenten in romanischen Sprachen. Tübingen. http://hypermedia.ids-mannheim.de/call/public/sysgram.ansicht?v_typ=d&v_id=392&v_wort=abt%C3%B6nungspartikeln http://www.canoo.net/services/OnlineGrammar/Wort/Adverb/Partikel/Modalpartikel.html</p>	particle	see "description" for examples	R

MERLIN ANNOTATION SCHEME (VII)						
PRAGMATICS TAGS (EA2 - annotations available for core corpus)						
Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
politeness	overly direct language form	P_Pol_dir	<p>Politeness is an attempt by speakers to avoid face-threatening acts or to soften them (Brown/Levinson 1975). More than defining this an error tag, we ought to speak about a lack of an appropriate level of politeness caused by too much directness. There are many reasons for politeness rules being transgressed, but here we only focus on directness.</p> <p>In the learner texts, the authors often tend to be overly direct, for example when they ask for information or want to complain about something. Involuntarily or not, the impression of directness may result from a vast variety of language forms, of course (e.g. lexical choices, punctuation).</p> <p>Here just as in S_Form_gen, it is up to the annotator to weigh his/her assessment of the appropriateness of the learner's language directness against what is required by the task and what is common in the target language.</p> <p>RELATED TAGS S_form_gen for specific language forms</p>	whole expression	<p>GER: "[Wir gehen ins Schwimmbad und du bringst etwas zu essen mit!]"</p> <p>CZE: "Teď mi řekni, co ti mám koupit za dárek."</p> <p>ITA: "[Senti, io penso a andare a trovarti, per vedere la tua città famosa.]" [Rispondimi presto, devo prenotare il volo.]</p>	T
REQUEST	direct REQUEST	P_Request_direct	<p>DESCRIPTIVE 'EXISTENCE TAG'</p> <p>All direct REQUESTS are tagged no matter if they contain errors. Errors get appropriate error tags in addition to the P_Request_direct tag. For a definition of the speech act, consult the glossary.</p> <p>Exemplary realization strategies for direct REQUESTS</p> <p>Direct requests are explicitly marked, for example by the use of</p> <ul style="list-style-type: none"> - imperatives, - mood derivatives [Stop talking. Re-write that paragraph.], - performatives [I'm asking you to rewrite the paragraph] - hedged performatives [I would like to ask you...], in German e.g.: [denk dran] [sei so nett] - obligation statements [You'll have to re-write that paragraph], - want statements [I really wish you'd stop chattering]. <p>Distinguishing REQUESTS from similar speech acts in MERLIN</p> <p>In MERLIN, the following speech acts are not tagged as REQUESTS:</p> <ul style="list-style-type: none"> - SUGGESTION (benefit for both the speaker and the hearer), - ADVICE/INSTRUCTION (benefit for only the hearer). (Source: Blum-Kulka/Olshtain 1984, Spencer-Oatley 2008). - pure REQUESTS for information ('Wie spät ist es?'). Only REQUESTS that demand an action by the hearer are tagged. <p>Range of REQUESTS (tag span)</p> <p>While the head act is always visible the other parts of the REQUEST need not necessarily be realized by the learner.</p> <p>a) Alerter, can be: address term(s) [Marie, ... excuse me etc.]</p> <p>b) Head act</p> <p>c) Adjunct(s) to head (i.e. preparator [I'd like to ask you something], Getting a precommitment [Could you do me a favour?], grounder [I missed class yesterday. Could I borrow your notes?], disarmer [I know you don't like to lend out your notes, but...], promise of reward [Could you give me a lift home? I'll give you sth for the petrol.], imposition downgraders [Could you lend me this book, if you're not using it at present?])</p>	(Alerter), head act, (Adjuncts) of the REQUEST (including all parts the learner realized linguistically)	<p>GER:</p> <p>Imperatives:</p> <ol style="list-style-type: none"> 1. Fragen Sie Ihre Julia. 2. Ruf mich bitte an 3. Bitte schreib mir zurück. 4. Bitte rufen Sie mir am Abend von 18 bis 22 Uhr an. 5. schick mir ein E-Mail oder ruf mich <p>Other:</p> <ol style="list-style-type: none"> 1. Aus ausgegebenen Gründe fördere ich mich Zurückerstattung diese Kosten. 2. Um meine „Wunden“ zu behandeln Bitte ich Sie um eine Entschädigung. 3. Ich möchte dich fragen vielleicht kannst du meine Fahrkarte benutzen, weil du schon am Wochenende deine Eltern in München besuchen wolltest 4. Bitte nicht vergessen! <p>CZE:</p> <p>*Prosim, poslej mně to inzerát. Mam jenom dvě nebo tři otázky... *Neříkáš jestli oslavy bude poledne nebo večer. Čekám za odpovědani.</p> <p>ITA:</p> <p>*Vi prego di farmi sapere se avete bisogno delle informazione ulteriore; *Fatemi sapere!; *Portarmi il libro!; *Dovresti pulire la cucina.; *Vorrei che tu mi accompagnassi alla stazione.</p> <p>Imperatives:</p> <p>Ne pensi e fammi sapere la tua decisione. Mi chiami per dirmi. Scrivimi una e-mail si tu hai voglia e tempo per vederci Se ancora avrei problemi, mi scriverai e proverò aiutarti di più</p> <p>Other:</p> <p>Le preghiamo compensazione.</p> <p>A questo punto la richiesta di un parziale rimborso del soggiorno trascorso</p>	R, T
REQUEST	indirect REQUEST	P_Request_indirect	<p>DESCRIPTIVE 'EXISTENCE TAG'</p> <p>All indirect REQUESTS are tagged regardless of errors. Errors get appropriate error tags in addition to the P_Request_indirect tag. For a definition of this speech act, consult the glossary.</p> <p>Indirect REQUESTS: exemplary realization strategies</p> <ul style="list-style-type: none"> - suggestive formula [How about re-writing that paragraph?] - query preparator, i.e. reference to preparatory conditions of the speech act [Could you stop talking, please?] - partial reference to object or element needed for the implementation of the speech act ('Why is the window open'), - reliance on contextual clues ('It's cold in here'). Hints can be stronger [You've made a lot of mistakes in that paragraph] or less strong [I'm getting a headache. --> context: other people are talking loudly]. <p>Distinguishing REQUESTS from similar speech acts in MERLIN</p> <p>In MERLIN, the following speech acts are not tagged as REQUESTS:</p> <ul style="list-style-type: none"> - SUGGESTION (benefit for both the speaker and the hearer), - ADVICE/INSTRUCTION (benefit for only the hearer). (Source: Blum-Kulka/Olshtain 1984, Spencer-Oatley 2008). - pure REQUESTS for information ('Wie spät ist es?'). Only REQUESTS that demand an action by the hearer are tagged. <p>Range of REQUESTS (tag span)</p> <p>While the head act is always visible the other parts of the REQUEST need not necessarily be realized by the learner.</p> <p>a) Alerter, can be: address term(s) [Marie, ... excuse me etc.]</p> <p>b) Head act</p> <p>c) Adjunct(s) to head (i.e. preparator [I'd like to ask you something], Getting a precommitment [Could you do me a favour?], grounder [I missed class yesterday. Could I borrow your notes?], disarmer [I know you don't like to lend out your notes, but...], promise of reward [Could you give me a lift home? I'll give you sth for the petrol.], imposition downgraders [Could you lend me this book, if you're not using it at present?])</p>	(Alerter), head act, (Adjuncts) of the REQUEST (including all parts the learner realized linguistically)	<p>GER: *Entschuldigung. Aber möchte ich ein PostCard von Istanbul kannst du mir schinken?, *Ich wünsche mir aus Istanbul einige Postkarte.</p> <p>*Können Sie bitte meine Katze füttern [?]</p> <p>*Können Sie bitte mir erklären, vielleicht mit einer Beispiel, was für Rechte und Pflichten einen normalen Au-pair habt</p> <p>*ich wollte dich fragen ich lasse ihn bei dir weil er braucht streichen und spatsseren mit ihn</p> <p>*Ich wäre Ihnen sehr dankbar, wenn Sie mir weitere Informationen über Rechte, Pflichten und Aufenthaltsformalitäten zuschicken könnten</p> <p>*Ich warte ungeduldig auf deine positive Antwort.</p> <p>*Lieber Team des Aupaiglobal ich warte dass Ihr mir alle informationen uber eure Angebote schickt und ich Danke euch für das Verständnis</p> <p>CZE: *Mohl bys mi poslat ten inzerát?, *Chtela bych uvidě-t tvůj novy byt! Hodi si ti to?</p> <p>ITA: *Potreste mandarmi la lista dei corsi al mare e la possibilità di alloggio</p> <p>* Puoi dirmi dove lavori adesso e che cosa fai?</p> <p>* Potreste organizzare un posto dove posso dormire?</p> <p>* Non è che mi potresti aiutare a spostare l'armadio?</p> <p>* Cosa dici, mercoledì sera?</p> <p>* È possibile di andare ancor' una volta al bar 'Pino' per bere una birra con i ragazzi?</p> <p>* Ti andrebbe di invitarmi a partire del lunedì.</p> <p>* Se decidi che non voi vedere più i ragazzi, meglio se cambi il tuo numero di cellulare.</p> <p>GERMAN and ITALIAN examples from Schönfelder 2013 (available online).</p>	R, T

MERLIN ANNOTATION SCHEME (I)

GLOSSARY

General remarks

columns in annotation scheme	in the "examples" column, only the tag in focus is put in brackets [xyz] and corrected if necessary {xyz}; other errors that would receive other MERLIN tags are not corrected to avoid confusion
[]	The square brackets in the example column contain the tag-relevant extracts of learner language expressions that deviate from the target language as tagged according to the MERLIN annotation scheme.
{}	The curly brackets in the example column contain a correction of the erroneous learner expression. It is only given if the erroneous learner form (in square brackets) is not self-explaining or ambiguous.
*	The asterisk is used before learner language expressions that deviate from the target language
tagspan	This term defines the boundaries of each annotation tag. In the column "tagspan" of the annotation scheme, the units inside the tagspan are defined. The minimum tagspan is 1 token.

Definitions

article	An article is a word that is used with a noun to indicate the type of reference being made by the noun. Articles can, for example, specify the grammatical definiteness of the noun; in German, articles are "der", "die", "das", "ein(e, er). In Italian "il", "la", "lo", "le", "li", "i", "gli", "un(a, o)", "del", "dello", "della", "dei", "degli", "delle"
	Other word categories such as pronouns (mein;dein;jede/r/s; diese/e/r) as well as the determiner 'etwas' do also count as articles here.
conjunction	A conjunction is a part of speech that connects two words, sentences, phrases or clauses together and expresses a logical or grammatical relation between the related elements. In general, a conjunction is an invariable grammatical particle, and it establishes coordinating or subordinating relations. (Bußmann, Lexikon der Sprachwissenschaft, 3. Aufl. (2002), Kürschner, Grammatisches Kompendium, 4.Aufl. (2003))
clitic	Clitics are: mi, ti, le, gli, ci, vi, loro, lo, la, li, le, si, ne
formulaic sequences	"[...] a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray 2002: 9).
grapheme	Definition: A grapheme is the smallest semantically distinguishing unit in a written language, analogous to the phonemes of spoken languages. A grapheme may or may not carry meaning by itself, and may or may not correspond to a single phoneme.
	A grapheme may include more than one letter (ex. <sch> or <ch> in German).
POS	Words in a language can be divided according to different grammatical characteristics. The ones which share similar properties belong to the same part of speech (POS) i.e., word class. In all our languages (GER, ITA, CZE) we find POS such as article, noun, verb, adverb, preposition, etc. (see canoo.net for GER; e.g. http://bara.ujc.cas.cz/psjc/search.php or http://ssjc.ujc.cas.cz/search.php for Czech)
phrase	A phrase is considered a unit of one or more words that build a functional unit, that are adjacent and that can be permuted/moved as a whole within a sentence. Each phrase has a head that determines the formal features of the other elements of the unit and that determines the name of the unit. You may consider the following phrase types:

	<ul style="list-style-type: none"> • Nominal phrase (NP; it includes also cases where a pronoun is the head): Das alte Haus (gehört ...). / Das Buch auf dem Tisch (gehört ...) / Sie (ist Ärztin). • Prepositional phrase (PP): (Wir treffen uns) am alten Marktplatz. // (Sie kümmert sich) um die außerordentlich ängstlichen Patienten ihrer Praxis. • Adverb phrase (AdvP): (Sie ist immer (hilfsbereit)). // (Das Buch befindet sich) ganz oben. • Adjective phrase (AdjP): (Sie ist immer) sehr hilfsbereit. / (Er ist) arm an Mitteln.
preposition	Prepositions are a grammatically distinct class of words which often express spatial or temporal relations. As the primary function is relational, a preposition typically combines with another constituent (called its complement) to form a prepositional phrase, relating the complement to the context in which the phrase occurs. (Huddleston, Rodney, and Geoffrey K. Pullum. (2002) The Cambridge Grammar of the English Language)
valency	Verb valency refers to the number of arguments controlled by a verbal predicate. Verb valency includes all obligatory arguments, including the subject of the verb. A complement can be realized as adjective phrase (*Die Sitzung dauerte lange.), adjunct phrase (*Er fühlte sich, als ob er allein auf der Welt sei.), adverbial phrase (*Er isst gut und gern.), nominal phrase (*Der Vertreter hat sich nicht mehr gemeldet. Ich suche den Mann mit dem Sonnenschirm.), prepositional phrase (*Ich denke oft an dich.), prepositional adverb (*Ich denke oft daran.), pronominal phrase (*Er hat ihn gesucht.) and subjunctive phrase (*Ich weiß nicht, ob ich ihm das gesagt habe.)
connector	Connectors are inflectible expressions that do not allocate case attributes to their syntactical environment. Connectors connect two expressions semantically. More details about connectors can be found in Pasch (2009:1).
lexical error	Definition 1: "A linguistic form, ... which, in the same context would in all likelihood not be produced by the learner's native speaker counterparts." (Lennon 1991, 182). Definition 2: "When the native speaker estimates tend towards a zero probability of usage [...] we say that the learners are exhibiting a negative linguistic property. Negative properties are those that fall outside the set generated and that are judged ill-formed by native speakers. It is generally desirable for learner to reduce their production containing this negative properties, or 'errors', in the interests of successful communication." (Hawkins & Buttery 2010: 4)
Sociolinguistic competence	Definition: the "capacity to recognize and produce socially appropriate speech in context" (Lyster, 1994: 263)
Abtönungspartikeln	Abtönungspartikeln are typical for discourse. They aim at expectations and attitudes of the speaker/hearer and help to integrate an utterance into the context. They occur in the German Mittelfeld, they cannot be asked for, and they do not form phrases. They can be combined ("Dann geh [halt doch] heut abend dahin!")

Speech act REQUEST	<p>A REQUEST is an illocutionary act whereby a speaker conveys to a hearer that he/she wants the hearer to perform an act which is primarily for the benefit of the speaker (can be request for an object, an action or some kind of service). REQUESTS can be direct or indirect. In a direct REQUEST, the speaker explicitly says what he means. No contextual information is needed. The REQUEST is syntactically marked as such.</p> <p>Indirect strategies are procedures that realize the act by reference to contextual preconditions necessary for its performance, as conventionalized in the three languages (in Searle's terminology: indirect speech acts). "The simplest cases of meaning are those in which the speaker utters a sentence and means exactly and literally what he says. (...) But, notoriously, not all cases of meaning are this simple; In hints, insinuations, irony, and metaphor (...) the speaker's utterance meaning and the sentence meaning come apart in various ways" (Searle 1993: 30). A very well-known example is "Could you pass me the salt, please", where what is being asked for is not someone's ability to pass the salt, but the speaker requests the hearer to do something indirectly.</p>
---------------------------	---