

MERLIN ANNOTATION SCHEME (I)
GRAMMAR TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
word order	word order in main clause	G_Wo_womaincl	<p>This tag is used if an error occurs in a main clause in the following cases:</p> <p>a) verb in the wrong position b) wrong order of syntactical items such as attribute, subject, object.</p> <p>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.</p> <p>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!)</p> <p>RELATED TAGS If a participle, auxiliary, or copula is in the wrong position, it is tagged as G_Verb_compl_Pos.</p>	wrong clause including punctuation mark	<p>a) *[Včera učil jsem se český.], *[Vielleicht du könntest mir bei meine Wohnungssuche helfen.], *[Sollst du Wasser und Bikini mitbringen.]; *[Per trovarlo ti vorrei aiutare.]</p> <p>b) *[Známe film český.] *[Učil se jsem český.], *[Könntest du bitte mich anrufen?], *[Köntest Du bitte mir helfen bei der Wohnungssuche?]; *[Mia famiglia sta anche bene.], *[Vorrei andare a ti trovare.]</p>	R/T/U
word order	word order in subordinate clause	G_Wo_wosubcl	<p>This tag is used if an error occurs in a subordinate clause in the following cases:</p> <p>a) verb in the wrong position b) wrong order of syntactical items such as attribute, subject, object.</p> <p>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.</p>	wrong clause including punctuation mark	<p>a) První je těžký znát, [co učít 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?]</p> <p>b) Maminka myslí, [že musí pracovat moc pěkně pro lepší budoucnost.] *Ich habe nicht verstanden, [von welche fährt der Zug Bahnhof.]</p>	R/T/U
negation	general negation error	G_Neg_neggen_Pos G_Neg_neggen_O G_Neg_neggen_Ch G_Neg_neggen_Ad	<p>This tag is used in case of</p> <p>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)</p>	<p>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</p>	<p>a) *[mám ne] kávu, *[půjdu neráno], *Ich habe Hunger [keinen]; *Io credo [non].</p> <p>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}.</p> <p>c) *Bohužel, nemám [ne] čas; *Ich habe [nicht] Zeit. *Er wird dort arbeiten [nein]; *{Non}, viene più tardi.</p> <p>d) *Man kann nicht auf solchen grössen Teil seiner Persönlichkeit zu verzichten, ohne seine psychische Gesundheit [nicht] zu schaden.</p>	R
negation	double negation	G_Neg_negdoub	<p>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.</p>	1 token (verb, existing part of the double negation)	<p>*[mám] žádný čas {nemám žádný čas}</p> <p>*nikdo [volal] → {nikdo nevolal}</p> <p>*nemám [nějaký] čas {nemám žádný čas }</p>	R

MERLIN ANNOTATION SCHEME (I)

GRAMMAR TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	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	<p>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:</p> <p>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).</p> <p>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.</p> <p>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.</p> <p>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.</p>	wrong argument (usually 1 token) or whole clause including punctuation mark (if argument is missing)	<p>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}."</p> <p>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.</p>	R
Agreement (subject and verb)		G_Agr	<p>An agreement error in MERLIN includes wrong grammatical forms in the following combinations of subject and verb (incl. copulative predicates):</p> <p>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</p> <p>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.</p> <p>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)</p>	wrong form (usually 1 token)	<p>1) 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</p>	R/C
reflexivity	reflexive pronoun	G_Refl_pronrefl_O G_Refl_pronrefl_Ad G_Refl_pronrefl_Ch G_Refl_pronrefl_Pos	<p>Errors in the use of reflexive pronouns in MERLIN are tagged in case of:</p> <p>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)</p> <p>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.</p>	wrongly positioned reflexive pronoun or token containing the reflexive pronoun, in case of missing reflexive annotate the verb (verb	<p>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] 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</p>	R/T

MERLIN ANNOTATION SCHEME (1)

GRAMMAR TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
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
inexistent inflection	adjective inflection	G_Inflect_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] {cizí} nabídkou	R
inexistent inflection	noun inflection	G_Inflect_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_Inflect_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_Inflect_verb and G_Inflect_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 [mir] {mir}; * welche Rechte und [Pflichte]... 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 [tícíech] {tisících} nápadů	R
inexistent inflection	verb inflection	G_Inflect_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] {přejete}	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

MERLIN ANNOTATION SCHEME (I)
GRAMMAR TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
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, *[il] {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 [knihovnici] {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_Inflect_noun .	wrong token	ITA: *Con riferimento [al] {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
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, *zítra [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 gezeigt] 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, 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

MERLIN ANNOTATION SCHEME (I)

GRAMMAR TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
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] gearbeitet; *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} beantworten 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 demonstruje	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_Comp_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
preposition		G_Prep_O G_Prep_Ad G_Prep_Ch G_Prep_Pos G_Prep_Merge G_Prep_Split	This tag is applied if: 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) 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. 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	1 token; in case of missing prep. --> phrase	a) *myslím [maminku] (na maminku), *dívá se [televizi]; *ich warte [Antwort], *dann warte ich [dich]; *mi da [possibilitá] 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) *sedl 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}	R/U
article		G_Art_O G_Art_Ch G_Art_Ad G_Art_Merge G_Art_Split	This tag is used in the following cases: 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) 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. RELATED TAGS If an article is involved, G_Art_Merge and G_Art_Split are used, not G_Prep_Merge.	1 or 2 tokens (1 token: wrong article, 2 tokens: prep + art not contracted) missing article: phrase	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*knk..., *[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}, *[Fr]alle {fra le} e) *[su l']erba {sull'}, *da la {dalla}	T

MERLIN ANNOTATION SCHEME (I)

GRAMMAR TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	EXAMPLES	SOURCE (R=research; U=users; C=CEFR; T=text analysis)
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>	<p>1 token;</p> <p>If conjunction is missing --> clause supposed to contain the missing element</p>	<p>a)</p> <p>*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; *lo studente, sta in classe {,che sta...}/*řekl mu nedělal to {řekl mu, aby to nedělal}</p> <p>b)</p> <p>*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)</p> <p>*řekl, [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; *sarà meglio [quando] uno di voi mi chiama</p> <p>d)</p> <p>nepíšeš, nevadí mi to ale</p>	R
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)</p> <p>b) missing clitic (G_Clit_O)</p> <p>c) missing pronoun particle/pronoun particle+verb (G_Clit_O)</p> <p>d) pronoun particle not required (G_Clit_Ad)</p> <p>RELATED TAGS</p> <p>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.</p> <p>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>	<p>a), b), d), token with the wrong, added or wrongly placed element</p> <p>c) verb</p>	<p>a) puoi [chiamarla] {puoi chiamarmi} ; vorrei [mandarla] il mio curriculum {vorrei mandarla 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>	<p>one token (wrong POS)</p>	<p>ITA:</p> <p>*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:</p> <p>*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:</p> <p>*bylo to pro mě [potěšený] {potěšení}, *musíš myslet [rozumný] {rozumně}, zdravý život bez [nemocní] {nemoci}, doufám, že počasí bude [dobře] {dobře}</p>	R

MERLIN ANNOTATION SCHEME (1)

ORTHOGRAPHY TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	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}; *viaggiare {viaggiare}; *nove {nuove} b) *Peter {Petr}; *čtvrtek {čtvrtek}; *purst {prst}; *červenec {červenci}; *banek {bank}; *schpazieren; *eigentlich {eigentlich}; *mann {man}; *Monat {Monat}; *reggionale {regionale}; *occasione {occasione} c) *svadba {svatba}; *těším se na [ji] {ni}; *schreipt {schreibt}; *volte {wolte}; *Auslander {Ausländer}; *soziologia {sociologia}, esperienza {esperienza}; *mo {ma} Examples for grammatical errors (not annotated as orthographical errors): *piše 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}; *[duvod] {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}; *štedrý {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 perché; * era sereno. [poi] è nevicato. d) *[FahrkarTe]; *ich [wollTe] ...	U

MERLIN ANNOTATION SCHEME (I)
ORTHOGRAPHY TAGS

Specification (1)	Specification (2)	Tag denomination	DESCRIPTION	TAGSPAN	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>*sM.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 [,] sorveva in un palazzo molto bello. {La scuola che ho frequentato per due anni a New York e che ora non esiste più sorveva 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 (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	<p>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"</p> <p>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.</p>
conjunction	<p>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))</p>
clitic	Clitics are: mi, ti, le, gli, ci, vi, loro, lo, la, li, le, si, ne
formulaic sequences	
grapheme	<p>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.</p> <p>A grapheme may include more than one letter (ex. <sch> or <ch> in German).</p>
POS	<p>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.</p> <p>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)</p>
phrase	<p>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:</p>

MERLIN ANNOTATION SCHEME (I)

GLOSSARY

	<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.)

Abbreviations in tag denominations (referring to the type of target language modification)

add	superfluous (added) element
ambig	ambiguous - type of error can't be specified
ch	wrong choice of element
merge	elements are wrongly merged
o	omitted element
pos	wrong position
split	elements are wrongly split



Literature