Multilingual Platform for the European Reference Levels: Interlanguage Exploration in Context

# Documentation of additional annotation decisions – for Italian

Please cite as: MERLIN project, Documentation of additional annotation decision for Italian, 2014, http://merlin-platform.eu

Teresa Knittel

# Annotation manual: Annotation decisions for Italian

This paper documents the decisions taken during the annotation process in cases of problems and doubts for the Italian texts. Only decisions that are not included in the Annotation Scheme are documented here. Technical issues are not part of this document.

### Explicit subjects (TH1, EA1)

As there are no clear rules for Italian, when the use of an explicit subject is an error (to be annotated) and when a stylistic matter (not to be annotated), we rely on the native-speaker competence for the decision when to correct and tag an explicit subject as valency error (G_Valency_complnumb_Ad). We try to intervene as little as possible, but correct sentences that are highly non-native like.

### Missing agreement between implicit subject and verb (TH1, EA1, TH2, EA2)

If the missing agreement between an implicit subject and the verb is only inferable from the context and the sentence in isolation is grammatically correct, we do not correct the verb in TH1. We correct it in TH2 and tag it with C_Coh_ref in EA2 (i.e. no G_Agr or S_Form_addr).

> *E finalmente, quando [vorrei] Iniziare?* → TH1: *...vorrei iniziare?* → TH2: *...vuoi iniziare?*
> EA2: C_Coh_ref

> *Come stai? Che cosa fai? [Ha] molto lavoro?* → TH1: *...Ha molto lavoro?* → TH2: *...Hai molto lavoro?*          EA2: C_Coh_ref

### Agreement errors in past participle (EA1)

We use G_Agr for wrong agreement between subject and past participle, and G_Verb_compl_Ch for wrong agreement between object and past participle. In addition, we use the tag G_Clit_xxx if the predicate involves a wrong clitic.

> *Dove ci sono [scritto] le cose che...* → TH1: *Dove ci sono scritte le cose che...*     EA1: G_Agr

> *Noi [le abbiamo aiutato]* → TH1: *Noi l'abbiamo aiutata*          EA1: G_Verb_compl_Ch

### Ambiguous learner errors (EA1)

In some cases the learner error is ambiguous, i.e. there are two or more possible interpretations of the error and therefore two or more possible tags could be applied. Usually we decide case by case. Here the more general decisions are documented.

*Orthographic error vs. lexical error*

Minor orthographic errors leading to the use of non-existing words or to a semantic error are tagged as orthographic errors and not as lexical ones (e.g. V_form_word_nonexist or V_semdenot_word).

1

Teresa Knittel

*mo* → TH1: *ma*          EA1: O_Graph_graphgen_Ch

*nove* → TH1: *nuove*          EA1: O_Graph_graphgen_O

*invio* → TH1: *invito*          EA1: O_Graph_graphgen_O

*e* → TH1: *è*          EA1: O_Graph_act_O

*a* → TH1: *ha*          EA1: O_Graph_graphgen_O


*Orthographic error vs. grammatical error (function words)*

If the learner uses a wrong function word, we opt for grammatical error, even if only one letter is wrong compared to the right function word.

*la* → TH1: *le*          EA1: G_Morphol_numb_wrong (article)

*al* → TH1: *dal*          EA1: G_Prep_Ch

*da* → TH1: *di*          EA1: G_Prep_Ch


*Orthographic error vs. grammatical error (G_Inflect_xxx_inexist) vs. lexical error (V_form_word_nonexist)*

We opt for G_Inflect_xxx_inexist, if the form used by the learner does not exist and the error is found in the inflectional ending.

If the inflectional ending is partly or completely missing, we also use G_Inflect_xxx_inexist.

*le [\*nazionalitè]* → TH1: *le nazionalità*          EA1: G_Inflect_noun_inexist

*la [\*nazion]* → TH1: *la nazione*          EA1: G_Inflect_noun_inexist


If the error is in the word stem, we opt for orthographic error (see also above).

*\*mageremo* →TH1: *mangeremo*          EA1: O_Graph_graphgen_O

*la [\*fabrica]* → TH1: *la fabbrica*          EA1: O_Graph_graphgen_O


**Grammatical errors in formulaic sequences** (TH1, EA1, TH2, EA2)

If there is at least one possible context in which the sentence is grammatically correct, we keep the learner version in TH1 and change it in TH2 to the version which is correct in the context. We tag it as form error in a formulaic sequence in EA2.

*Presto!* (at the end of a letter) → TH1: *Presto!* → TH2: *A presto!*          EA2: V_FS_Form_O

(Possible context: *Quando vieni? – Presto!*)

If there is no context in which the sentence is grammatically correct, we correct it in TH1 and tag it in EA1 with the appropriate grammatical error tags. In addition, we use the appropriate tag for form errors in formulaic sequences in EA2.

*giocare [al] calcio* → TH1: *giocare a calcio* → TH2: *giocare a calcio*        EA1: G_Art_Ad + EA2: V_FS_Form_Ad

*Mi piacerebbe frequentare [sul] questo corso.* → TH1: *...frequentare questo corso.* → TH2: *...frequentare questo corso.*        EA1: G_Prep_Ad + G_Art_Ad, EA2: V_FS_Form_Ad

## Wrong use of articles and partitive articles which can only be inferred from the context (TH1, EA1, TH2, EA2)

In cases of sentences in which the article or partitive article used is formally correct (i.e. the sentence is grammatically correct), but is not right in the global context, we correct the article or partitive article in TH2. No grammatical error tag can be applied in EA2, so the correction remains untagged.

*mangiare i panini* → TH2: *mangiare dei panini*        EA2: no tag

*frequentare i corsi di italiano* → TH2: *frequentare dei corsi di italiano*        EA2: no tag

## TH1 corrections that become incorrect in TH2 (TH1, EA1, TH2, EA2)

If a wrong verb in TH1 (that can be corrected only in TH2) requires another preposition as the one written by the learner, we correct the preposition in TH1 and tag it in EA1. In the second step (TH2) we correct the verb and add the preposition required by the new verb, but in EA2 we tag only the semantic error of the verb.

*partire [a] Milano* → TH1: *partire per Milano* → TH2: *{andare a} Milano*
EA1: G_Prep_Ch; EA2: V_semdenot_word_1

## Articulated prepositions (EA1)

We consider the componential character of articulated prepositions (preposition + article) and tag the element(s) of the articulated preposition which is (are) wrong (i.e. wrong preposition, wrong article).

*giocare [al] calcio* → TH1: *giocare a calcio*        EA1: G_Art_Ad (not G_Prep_Ch)

*frequentare [sul] questo corso.* → TH1: *frequentare questo corso.*        EA1: G_Prep_Ad + G_Art_Ad

## Wrong article contraction/Non-existing articulated prepositions (EA1)

Non-existing contractions of preposition and article are tagged with G_Inflect_noun_inexist.

*[\*nelli] programmi* → TH1: *nei programmi*        EA1: G_Inflect_noun_inexist

Ambiguous contractions like *"nell Corriere"* are tagged as orthographic errors.

*[nell] Corriere* → TH1: *nel Corriere*        EA1: O_Graph_graphgen_Ad

3

Teresa Knittel

**Wrong/non-existing gender of nouns (EA1)**

If the learner uses a wrong, non-existing gender of a noun, we tag it with G_Inflect_noun_inexist, not with G_Morphol_gend_wrong. (G_Morphol_xxx_wrong is only used for forms that exist in the inflectional paradigm.)

We don't use V_form_word_nonexist, as the main problem is not that the word does not exist, but that the learner has chosen a wrong gender. So we tag a grammatical error.

| | |
|---|---|
| *un'aiuta* → TH1: *un aiuto* | EA1: G_Inflect_noun_inexist |
| *gruppa* → TH1: *gruppo* | EA1: G_Inflect_noun_inexist |

If the gender error of the noun leads to the use of an existing form of a different part of speech (e.g. verb form, adjective), but it is clear from syntax that the learner used the word as noun, we use G_Inflect_noun_inexist and not G_POS.

| | |
|---|---|
| *ho molto [lavora]* → TH1: *ho molto lavoro* | EA1: G_Inflect_noun_inexist |
| *fare [passeggiati]* → TH1: *fare passeggiate* | EA1: G_Inflect_noun_inexist |

**Concordance (EA1)**

If the learner uses the wrong gender for a noun and uses the same gender for article, prepositions ecc., only the gender error of the noun is tagged, because the learner is coherent in the choice of the gender.

| | |
|---|---|
| *un'aiuta* → TH1: *un aiuto* | EA1: G_Inflect_noun_inexist for noun, no error tag for article |

**Wrong use of interrogative and demonstrative adjectives (TH1, EA1, TH2, EA2)**

If interrogative adjectives are used instead of demonstrative adjectives and viceverse, we tag it as semantic error in EA2, as they belong to the same part of speech and thus we can't apply G_POS. There is also no other appropriate grammatical tag. The error is therefore corrected in TH2 and not in TH1.

*Puoi mi dire dove lavori adesso in quale ufficio e in [quella] città.* → TH1: *...e in quella città?* → TH2: *...e in quale città?*          EA2: V_semdenot_word_1

*Vorrei sapere [quelle] parti ti piacciono.* → TH1: *...quelle parti...* → TH2: *...quali parti...*          EA2: V_semdenot_word_1

*Puoi dirmi a [questa] ditta hai lavorato in tempo scorso.* → TH1: *...in questa ditta...* → TH2: *...in quale ditta...*          EA2: V_semdenot_word_1

**Clitics (EA1, EA2)**

If a learner omits a clitic and therefore an obligatory object, we tag it with both tags: G_Clit_O (clitic is missing) and G_Valency_complnumb_O (obligatory object is missing).

Teresa Knittel

Every time that a clitic is missing or is wrong, it also involves a reference problem. Therefore we additionally use C_Coh_ref in EA2.

> *Ho dimenticato di [scrivere] prima.* → TH1: *...di scriverlo prima.*
> EA1: G_Clit_O + G_Valency_complnumb_O; EA2: C_Coh_ref

If the learner uses a wrong clitic, we use G_Clit_Ch and not G_Morphol_xxx_wrong (the latter tag includes pronouns and clitics are pronouns), because G_Clit_Ch is the more specific tag. Like above, we additionally use C_Coh_ref.

> *puoi [chiamarle]* → TH1: *puoi chiamarla*          EA1: G_Clit_Ch; EA2: C_Coh_ref


**Errors in double pronouns (EA1)**

We use G_Clit_Ch for tagging errors in double pronouns.

> [*vi] la possiamo regalare* → TH1: *ve la possiamo regalare*          EA1: G_Clit_Ch


**Negation (TH1, EA1)**

Missing negation words like in the following example are added in TH1 and tagged as G_Neg_neggen_O.

> *non fare [qualche cosa] di...* → TH1: *non fare niente di...*          EA1: G_Neg_neggen_O
> (tagspan: [qualche cosa])


**Apostrophes and Word boundaries (EA1)**

If an apostrophe is missing and both words are attached to one another, we use O_Apostr_O and O_Wordbd_Merge.

> *\*daccordo* → TH1: *d'accordo*          EA1: O_Apostr_O + O_Wordbd_Merge
>
> *\*anchio* → TH1: *anch'io*          EA1: O_Apostr_O + O_Wordbd_Merge


**Word boundaries: General tag O_Wordbd_Merge/Split and more specific tags G_Prep_Merge/Split and G_Art_Merge/Split (EA1)**

O_Wordbd_Merge/Split is the general, overarching tag for all merges/splits, whereas G_Prep_Merge/Split and G_Art_Merge/Split are special cases of word boundary errors. That means that every G_Prep_Merge/Split and G_Art_Merge/Split is at the same time an O_Wordbd_Merge/Split. Therefore every word boundary error gets the tag O_Wordbd_Merge/Split and, if it involves prepositions or articles, G_Prep_Merge/Split or G_Art_Merge/Split has to be applied additionally.

In cases in which articles and prepositions are merged/splitted, we use G_Art_Merge/Split, not G_Prep_Merge/Split.

> [*perla] casa* → TH1: *per la casa*          EA1: O_Wordbd_Merge + G_Art_Merge

*rispondi mi!* → TH1: *rispondimi!*          EA1: O_Wordbd_Split

**Diacritics (TH1, EA1)**

If an accent is used instead of an apostrophe, we use two tags: O_Apostr_O and O_Graph_act_Ad.

*un [pò]* → TH1: *un po'*          EA1: O_Apostr_O + O_Graph_act_Ad

**Punctuation (TH1, EA1)**

We don't regard it as obligatory to use a comma between the greeting and the name of the author at the end of a letter / e-mail. So we don't add a comma in TH1 if the learner has not used one. Commas used by the learner are ok.

*A presto Maria* → TH1: *A presto Maria*          EA1: no tag

*A presto, Maria* → TH1: *A presto, Maria*          EA1: no tag

*E-mail* vs. *email*: We accept both forms (reference: Zingarelli 2013) and don't correct in TH1.

***Troncamento* and *elisione* (EA1)**

We tag cases of wrongly used *troncamento* or *elisione* as orthographic errors, as we don't have an appropriate grammatical tag for this (e.g. *troncamento* instead of *elisione*, *elisione* instead of *troncamento,* unnecessary *troncamento*).

*[quel] esame* → TH1: q*uell'esame*          EA1: O_Graph_graphgen_O + O_Apostr_O

*[bel'] lavoro* → TH1: *bel lavoro*          EA1: O_Apostr_Ad

*[qual] bisogno avete?* → TH1: *quale bisogno avete?*     EA1: O_Graph_graphgen_O

**Facultative elision (TH1, EA1)**

We correct and tag if the learner's version sounds very non-native-like. If the facultative elision used by the learner is acceptable we don't correct and don't tag.

*d'italiano* → TH1: *d'italiano*          EA1: no tag

*di italiano* → TH1: *di italiano*          EA1: no tag

*una impresa* → TH1: *una impresa*          EA1: no tag

*un'impresa* → TH1: *un'impresa*          EA1: no tag

Errors in obligatory elisions are tagged as orthographic errors (see above).

### *d eufonica* (TH1, EA1)

We only correct and tag if the learner's version sounds very cacophonous and thus non-native-like. In the other cases we leave the learner's version and don't tag. Tag to be used if correction is necessary: O_Graph_graphgen_O/Ad.

> *ed a fare* → TH1: *ed a fare*         EA1: no tag

### Capitalization of *signora/-e/-i* (TH1, EA1)

We accept lower and upper case, but the learner should use it coherently in the text. In case the learner uses both variants in one text, we correct to the form mostly used in the text and tag the corrected forms with O_Capit.

> *Gentili signori* → TH1: *Gentili signori*        EA1: no tag
>
> *Gentili Signori* → TH1: *Gentili Signori*        EA1: no tag

### Capitalization of personal pronouns and adjectives (TH1, EA1)

We accept lower and upper case, but the learner should use it coherently in the text. In case the learner uses both variants in one text, we correct to the form mostly used in the text and tag the corrected forms with O_Capit.

> *lei* & *Lei, suo* & *Suo, voi* & *Voi, vostro* & *Vostro ecc.*

### Capitalization of names of months (TH1, EA1)

We only accept lower case. We correct and tag both if the month is mentioned in the text and if it's in the beginning of the text, i.e. in the header.

> *5 [Agosto] 2010* → TH1: *5 agosto 2010*        EA1: O_Capit

### Errors in proper names (EA1)

We annotate errors in proper names only for:

a) names of countries, continents, oceans
b) names of Italian regions / federal states (endonyms)

Errors in other types of proper names (e.g. streets, companies, organizations) are corrected in TH1 but not tagged as error in EA1.

> *Nuova [Zealanda]* → TH1: *Nuova Zelanda*        EA1: O_Graph_graphgen_Ad
>
> *Via [Cavur]* → *TH1: Via Cavour*        EA1: no tag

.

.