**The Louvain Error Tagging Manual
Version 2.0**

Sylviane Granger, Helen Swallow and Jennifer Thewissen

2022

CECL Papers 4

**The Louvain Error Tagging Manual
Version 2.0**

Sylviane Granger, Helen Swallow and Jennifer Thewissen

Table of Contents

A. INTRODUCTION

This document is the fourth version of the Louvain Error Tagging Manual. The first version, version 1.1, was compiled in the 1990s at the Centre for English Corpus Linguistics (Université catholique de Louvain, Belgium) with a view to carrying out error analysis on the basis of learner corpora. This practice became known as computer-aided error analysis (see Dagneaux et al. (1998)[1] and Granger (2003)[2] for a detailed explanation of computer-aided error analysis). Although the Louvain Error Tagging Manual targets actual errors, it also includes a category that covers infelicities, i.e. non-erroneous but odd-sounding language (cf. the difference between errors and infelicities in Section B.2). The tagset was initially developed to tag the first version of the *International Corpus of Learner English* (ICLE) (Granger et al. 2002)[3] and further refined within the scope of the two subsequent versions of the ICLE (Granger et al. 2009[4] and 2020[5]). It has been used in many other learner corpus projects internationally.

The Louvain Error Tagging Manual is accompanied by a computer software program, the Université Catholique de Louvain Error Tagging Editor (UCLEE), which speeds up the process of inserting error codes and corrections into learner texts. A new version of the software (UCLEEv2) is currently under development[6]. In the examples included in this manual, error tags and corrections are coded in the format that is automatically assigned in UCLEE. An example of the coding used is <GNN corr="lives">*life*</GNN>, which marks up a noun number error (use of the singular *life* instead of the plural *lives*)[7]. Note that only the targeted error is corrected in the examples; the other errors are left uncorrected.

---

[1] Dagneaux E., Denness S. and Granger S. (1998) Computer-aided Error Analysis. *System: An International Journal of Educational Technology and Applied Linguistics* 26(2), 163-174.
[2] Granger S. (2003) Error-tagged learner corpora and CALL: a promising synergy. *CALICO* 20(3), 465-480.
[3] Granger S., Dagneaux E. and Meunier F. (2002) *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
[4] Granger, S., Dagneaux, E., Meunier, F., Paquot, M. (2009). *The International Corpus of Learner English. Handbook and CD-ROM. Version 2*. Louvain-la-Neuve: Presses universitaires de Louvain.
[5] Granger, S., Dupont, M., Meunier, F., Naets, H. & Paquot, M. (2020) *The International Corpus of Learner English. Version 3*. Louvain-la-Neuve: Presses universitaires de Louvain. https://corpora.uclouvain.be/cecl/icle/trial/
[6] Centre for English Corpus Linguistics (forthcoming). *Université catholique de Louvain Error Editor. Version 2.0*. Louvain-la-Neuve: UCLouvain.
[7] In the previous versions of the manual and UCLEE, the error code was enclosed by brackets and placed directly before the erroneous word/string, and the correction was inserted between dollar signs after the error, e.g. (GNN) life $lives$.

This error tagging manual explains how and why a given tag is attached to a specific type of error. On the basis of a corpus of non-native essays, in this case the ICLE, different categories of error have been defined, and a code has been assigned to each of them. Errors have been classified descriptively. No attempt is made to identify the source of errors, but annotators are free to add 'F' at the end of the tag if they suspect that the error is due to interference from the learner's mother tongue (see B.5: Transfer-related errors).

Error categories must be clearly defined and overlap as little as possible. Version 3 of the tagging manual establishes seven main categories:

1. **Formal errors**
2. **Grammatical errors, i.e. errors that break general rules of English grammar**
3. **Lexico-grammatical errors, i.e. errors that contravene word grammar, namely cases where the lexico-grammatical properties of a word have been violated**
4. **Lexical errors, i.e. errors involving the semantic properties of single words or phrases**
5. **Word Redundant, Word Missing and Word Order errors**
6. **Punctuation errors**
7. **Infelicities**

Most of these categories are divided into sub-categories.

It must be emphasised that these category distinctions are not absolute. There is a certain amount of overlap, resulting in "fuzzy" areas. In order to prevent errors from being classified arbitrarily, a practical guide is provided at the end of the manual (cf. E: Fuzzy areas: Practical guide). Warning boxes have also been inserted at relevant points, alerting the annotator to exceptions. These should help annotators confronted with a fuzzy area to choose the tag that best corresponds to the error they wish to tag. However, in spite of these efforts, it is clear that error annotation will always contain an element of subjectivity. As pointed out by Milton & Chowdhury (1994: 129)[8], "[t]agging a learner corpus allows us, at least and at most, to systematize our intuitions".

There are 54 error tags (cf. D: Index). The first letter of the tag indicates the error category: F stands for Form, G for Grammar, X for LeXico-grammar, L for Lexis, etc. The second letter generally indicates the sub-category, and the letters that follow provide further precision about the

---

[8] Milton, J. & Chowdhury, N. (1994). Tagging the interlanguage of Chinese learners of English. In L. Flowerdew & A.K. Tong (eds) *Entering Text*. Hong Kong: The Hong Kong University of Science and Technology, 127-143.

type of error. Let us take, for example, the numerous tags that begin with G, denoting grammatical errors. A grammatical error may, for instance, involve a problem of subject-verb agreement, e.g. *Progress <u>have</u> been made*. In such a case the GVN tag will be used to indicate that the error pertains to grammar, and more specifically the <u>v</u>erb and its <u>n</u>umber.

The following point should be borne in mind when using the manual: the annotation system was originally intended for the analysis of learner writing rather than speech. Researchers who wish to use the manual to analyse speech will need to make the necessary changes (on this issue, see Gilquin & De Cock 2011[9]).

B.   IMPORTANT GUIDELINES

1. Tagging discrete errors vs rewriting

As the purpose of error tagging is primarily to uncover discrete errors, it is important that the annotator should try to identify these rather than **rewriting whole chunks of text**. Annotators should therefore try to stick as closely as they can to the learner's text, departing as little as possible from the original piece of writing. The following is an example of a grammatical error that would be overlooked as a result of rewriting:

When a thrilling program is on, we turn into sponges that imbibe all what is coming

This sentence should not initially be tagged as an infelicity and corrected as

....we passively absorb everything that comes onto the screen

Rewriting the original sentence would mean that the pronoun error (*all what*) went unnoticed. In cases like this, the annotator should make sure that they highlight the grammatical error in the pronoun (GP). However, if they feel it necessary, they can also mark the stylistic problem using the Z tag, thereby bringing into play the double-tagging system (cf. Principle 3). The annotator must resist the urge to immediately rewrite the student's original version, concentrating instead on actual errors, even if the corrected version still does not sound 100% native-like.

---

[9] Gilquin, G. & De Cock, S. (2011). Errors and disfluencies in spoken corpora: Setting the scene. *International Journal of Corpus Linguistics* 16(2),141-172.

2. Errors vs infelicities

The tagging manual makes an important distinction between real errors and infelicities. Real errors involve the breaking of a specific linguistic rule (grammatical, lexical or other), while infelicities are instances of non-erroneous, but odd-sounding, language (see examples below). Infelicities include:

1. **Register problems**, such as contractions (*couldn't*), informal expressions (*a bit, lots of*), slang, interjections, etc.
2. Questions of **political correctness** (*chair<u>man</u>* instead of the politically correct *chairperson*, *<u>man</u>kind* instead of *humankind*, *he/she, himself/herself*).
3. **Stylistic problems** (*What else could be the reason for this resistance…* instead of *What other reason could there be for this resistance…*).

Distinguishing between errors and infelicities in the error detection process should make for greater consistency within the category of errors proper. As shown by two pilot studies conducted at UCLouvain[10], the number of errors detected by two different scholars annotating the same texts can differ dramatically. Although differences between annotators are unavoidable, distinguishing carefully between real errors and infelicities will make for a higher rate of quantitative consistency in the real error category.

Infelicities are all tagged <Z> and are rephrased appropriately.

Examples of infelicities

The student has to take care of <Z corr="him- or herself">*himself*</Z> (political correctness)

This intimate place <Z corr="could not">*couldn't*</Z> *possibly be harmed by simple material conditions*. (Register problem)

It is all too easy to get guns illegally <Z corr="even if legal access is restricted">*when the legal way is blocked*</Z> (Stylistic problem)

To conclude, <Z corr="0">*I shall once more insist on the fact that*</Z> *a world from which dreams and imagination are absent would certainly not be worth living in.* (Stylistic problem: long-winded introductory sentence)

---

[10] These two pilot studies consisted mainly in two native speakers being asked to independently detect and correct errors found in the same texts. The results showed that the number of errors detected differed considerably from one annotator to the next, with one rater sometimes detecting twice as many errors in a text as their colleague.

3. Limited tagset

The tagset is limited: it comprises only 54 tags. At this stage this is felt to be an advantage: having to deal with too many tags would only make the annotator's job more difficult.

4. Error tagging principles

Principle 1

Do not tag on the basis of the corrected word/phrase, but on the basis of the incorrect word/phrase.

The main feature of a campus like Louvain-la-Neuve is <GA corr="its">*the*</GA> *conviviality.*

The error should be tagged <GA> (misuse of article) rather than <GDO> (misuse of a possessive determiner).

Exceptions

1. <GNC> is used to tag errors relating to the Saxon genitive. It is used to tag **both** erroneous use of the Saxon genitive **and** erroneous use of a phrase where a Saxon genitive would be appropriate.

- The Saxon genitive should not have been used:

    Behind the <GNC corr="Berlin wall">*Berlin's wall*</GNC>

    <GNC corr="Rehabilitation centres">*Rehabilitation's centres*</GNC>

- The Saxon genitive should have been used:

    <GNC corr="My sister's car">*The car of my sister*</GNC>

2. <GNN> is used to tag an *'s* or *s'* used wrongly as a plural ending:

The weak points of feminism centre on the fact that women will always feel like underprivileged <GNN corr="beings">*being's*</GNN>

Principle 2

Place the tag on the error (word or phrase) that needs to be corrected.

This type of <GNN corr="exchange">*exchanges*</GNN>

Exceptions

1. The <X*CO> and <X*PR> tags should be placed on both the word or phrase that <u>triggers</u> the erroneous complementation or preposition and the erroneous complementation or preposition itself. <X*CO> is used to tag lexico-grammatical errors involving complementation problems, while <X*PR>is used to tag lexico-grammatical errors involving dependent prepositions.

Students have the <XNCO corr="possibility of leaving">*possibility to leave*</XNCO>*.*

Besides, the latter are <XADJPR corr="dependent on">*dependent for*</XADJPR> *the former.*

2. The <XNUC> tag, which is used for errors that affect uncountable nouns**,** should be placed on both the incorrect article and the noun.

<XNUC corr="a piece of information">*an information*</XNUC>

Principle 3

When there are two types of error in the same word/phrase, double-tag the word/phrase.

 In the example below: first, use <LSN> to highlight the lexical error (*attraction*); second, use the <GNC> tag to highlight the genitive error (*landscape's attraction*).

the increasing number of factories has spoiled the <GNC corr="beauty of the landscape">*landscape's*<LSN corr="beauty">*attraction*</LSN></GNC>

Another example of double tagging: spelling error and count/uncount error

<XNUC corr="information"><FS corr="information">*informatoins*</FS></XNUC>

Principle 4

Use '0' (zero) to indicate a **missing word** or a correction that consists in **the deletion of a word/phrase**. Even if more than one word has been added or deleted, only **a single '0'** should be used (cf. Sections 5.1 and 5.2).

They are doing everything possible <WR corr="0">*they can*</WR> *to avoid the war.*

Principle 5

When the learner uses a word that exists in English but whose meaning in the context is unclear, the <LS> category should be used.

If the annotator is able to suggest a possible alternative, this should be inserted, followed by a question mark. **Only one suggestion** should be inserted, even when several spring to mind; the one most prominent in the annotator's mind should be chosen[11]:

Sometimes we cannot blame the masses for being <LSADJ corr="small-minded?">*narrow*</LSADJ>

If the annotator is unable to guess at the meaning, only a question mark should be inserted.

If we were to assume that technological progress and human imagination are two irreconcilable things, then we would <LSADV corr="?">*imminently*</LSADV> *be taking on a Romantic or even Luddite attitude.*

Principle 6

Do not tag errors resulting from the correction of a previous word.

Resorting to violence might seem like a <LSADJ corr="understandable">*comprehensible*</LSADJ> *reaction.*

---

[11] For a description of error annotation systems that allow for multiple corrections, see A. Lüdeling & H. Hirschmann (2015). Error annotations systems. In Granger, S., Gilquin, G. & Meunier, F. (eds) *The Cambridge Handbook of Learner Corpus Research*, CUP, 135-158.

Only tag the lexical error (*comprehensible* instead of *understandable*); do not tag the article issue (use of *a* instead of *an*) as the need to use *an* is a direct result of the correction, not an initial error by the learner.

Exceptions

Correcting a **tense error** may lead to the need for other tense changes even though in the original context there is no error. In such cases, all auxiliaries and tense forms that need changing should be tagged separately.

> *In the meantime the young man became a very likeable personality. He learned a profession in the prisson, and now he* <GVT corr="writes">*wrote*</GVT> *poetry and* <GVT corr="takes">*took*</GVT> *part in the publication of a prison's journal.*

5. Transfer-related errors

The error tagging system described in this manual is descriptive, not explanatory. Hence it does not make any suggestions as to the possible source(s) of the errors. However, annotators particularly interested in transfer-related errors may choose to add 'F' (Trans<u>F</u>er) at the end of a tag when they come across an error that they suspect of being due to interlingual interference.[12] For example, the suspected false friend error produced by a French learner *He *pretends to be ill* (FR *prétendre*) can be tagged *He* <LSVF corr="claims">*pretends*</LSVF> *to be ill* if the annotator wants to focus on cases of suspected interference errors at a later stage.

6. Checking the consistency of the tagging

Annotators should check the consistency of their tagging. This can be done, for example, by calling up a concordance of selected tags with a text retrieval software program such as WordSmith Tools[13] or AntConc[14] or within the UCLEE interface itself, as it includes a checking function and a concordancer. By doing this we became aware, for example, of several instances where <GA> was used instead of <XNUC>: *a device of such* <XNUC corr="0">*a*</XNUC> *questionable value*. One way to detect such cases is by carrying out a concordance of <GA> and sorting the

---

[12] In UCLEE, this possibility has already been implemented for lexical errors.
[13] Scott, M. (2020). *WordSmith Tools* Version 8. Stroud: Lexical Analysis Software.
[14] Anthony, L. (2022). *AntConc* Version 4.0.3. Tokyo: Waseda University.

concordance lines on first right (1R). This will give you all the instances of 'a' or 'an' and enable you to detect any mistagged segments. Similarly, by carrying out a concordance of all the <WM> tags, we detected cases where <WM> was used instead of <GPP> for missing personal pronouns: *Leontes sends one of his men to kill her but* <GPP corr="he">*0*</GPP> *doesn't.*

7. Annotators

To ensure that annotation is carried out effectively, it is essential that annotators are trained to error tag and that they familiarise themselves with the tagging manual. They should be native or expert users of English. It is an added advantage if they also have a good knowledge of the learner's mother tongue.

C. ERROR CATEGORIES

1. FORM <F>

This category encompasses two formal error categories: (1) spelling errors, which are tagged <FS>, and (2) morphological errors on derivational affixes, which are tagged <FM>.

1.1. Spelling <FS>

<FS> covers all **spelling errors**. It includes
- misuse or omission of capital letters:
    <FS corr="Flemish">*flemish*</FS>
    <FS corr="English">*english*</FS>
- issues concerning the doubling of consonants/vowels:
    <FS corr="aggressions">*agressions*</FS>
    <FS corr="altogether">*alltogether*</FS>
- letter substitution or misordering:
    <FS corr="unemployment">*unimployment*</FS>
    <FS corr="independence">*independance*</FS>

        `<FS corr="people">`*poeple*`</FS>`

- misuse/omission of hyphens/blanks between words or parts of words:

        `<FS corr="inhabitants">`*in-habitants*`</FS>`

        I don't need `<FS corr="anything">`*any thing*`</FS>` *else.*

- homophones (*it's/its, their/there*, etc.):

        `<FS corr="their">`*there*`</FS>` *houses*

        despite `<FS corr="its">`*it's*`</FS>` *efforts*

- slips of the pen which result in a spelling error:

        To start off with, we can `<FS corr="say">`*sy*`</FS>` *that this phrase has much to do with…*

Word coinages and unwarranted use of mother-tongue words or phrases where an English equivalent is available are also included in this category:

        He has been `<FS corr="dispossessed">`*aparted*`</FS>` *of his father's riches*

        Men do their `<FS corr="military service">`*servicio militar*`</FS>` *then need to find a job.*

1.2. Morphology <FM>

<FM> is used for **morphological errors**; it is limited to cases of erroneous <u>derivational</u> affixes. Errors involving <u>inflectional</u> morphology (*he has *swam, the *mouses, the *poors*) are classified in the corresponding grammatical error categories.

        `<FM corr="symbolic">`*symbolical*`</FM>`

        `<FM corr="impossible">`*unpossible*`</FM>`

        `<FM corr="inconceivable">`*unconceivable*`</FM>`

        `<FM corr="cowardice">`*cowardness*`</FM>`

        `<FM corr="talented">`*talentuous*`</FM>`

        `<FM corr="valuable">`*valious*`</FM>`

        `<FM corr="cultured">`*culturated*`</FM>`

2. GRAMMAR <G>

This category groups together errors that contravene general rules of English grammar. It is broken down into the following sub-categories:

2.1. Determiners <GD>

Determiners are words such as *my, this, some, each, any* or *which* used before nouns. Some of them (*this* or *some*) can also be used as pronouns, and when that is the case the error is tagged in the pronoun category <GP>. The <GD> category includes all determiners (demonstrative, possessive, indefinite, etc.), with the exception of articles, which are placed in a separate category (cf. Section 2.2).

2.1.1. Demonstrative determiners <GDD>

This category involves all errors affecting demonstrative determiners.

<GDD corr="These">This</GDD> elements cannot be separated.

They have largely lost <GDD corr="the">those</GDD> moral guidelines that…

I think that <GDD corr="the">this</GDD> Industrial Revolution should rather be seen as…

The first revolutionary ideas are forgotten. That happens even with <GDD corr="those">that</GDD> ideas which were good and fair.

They used <GDD corr="these">this</GDD> plays as social criticism.

2.1.2. Possessive determiners <GDO>

This category includes all errors pertaining to possessive determiners.

We see how women's role was changing: they have a strong character and it is <GDO corr="their">her</GDO> opinion which would determine the marriage contract.

Hermione invites Polixenes to stay at their home. Leontes thinks <GDO corr="his">her</GDO> wife is deceiving him with his friend.

People accept jobs according to how much they get paid but not according to <GDO corr="their">his</GDO> preferences.

2.1.3. Indefinite determiners <GDI>

This category involves all errors affecting indefinite determiners (i.e. *all, another, any, each, either, neither, most, both, few, many, much, more, several, some*).

He does not have <GDI corr="any">some</GDI> expectations.

There are too <GDI corr="many">much</GDI> students.

2.1.4. Other types of determiners <GDT>

This category includes all errors involving types of determiner other than those found in the

> **Warning!**
>
> Missing or redundant determiners are tagged <GD*>, not <WM> or <WR>.
>
> > *We all know that this will not change, until the people who are requiring* <GDO corr="0">*their*</GDO> *cheaper labour accept to raise salaries too.*
> >
> > *Women had come to believe staying at home was their only place in life and that they were not independent people with* <GDO corr="their">*0*</GDO> *own aspirations.*

first three categories (relative determiners, interrogative determiners, etc).

2.2. Articles <GA>

Although articles are a type of determiner (central determiners), they are tagged separately as they are particularly error-prone and therefore merit a category of their own.

The <GA> category covers all problems of definite, indefinite or zero articles where the article error contravenes <u>general rules of grammar</u>. This involves erroneous use of the article with definite, indefinite or generic reference.

Definite

> *People remember the dropping of* <GA corr="the">*0*</GA> *atomic bomb on Hiroshima.*
>
> *His symptoms seemed to suggest* <GA corr="a">*the*</GA> *fall into despair.*
>
> <GA corr="0">*The*</GA> *life is beautiful*
>
> *Total equality between* <GA corr="0">*the*</GA> *human beings nevertheless belongs to the realm of utopia.*

Indefinite

> <GA corr="a">*an*</GA> *utopia*

Zero

> *not only does* <GA corr="the">*0*</GA> *economy take over…*

In the case of <GA corr="a">*0*</GA> *handicap, the inequality is even more striking.*

Warning!

Note that article errors relating more specifically to <u>word grammar</u> are not tagged <GA>. These comprise: use of an indefinite determiner with an uncountable noun <XNUC>, and erroneous use of articles contained within connectors (tagged <LC*>) or multi-word phrases (tagged <LP>).

> <LCLC corr="On the one hand">*On one hand*</LCLC>, *nature started to deteriorate because man became a terrible predator.*

> *They could not tolerate that the biggest reserve in the world of 'Black Gold', by that time* <LP corr="in the hands of">*in hands of*</LP> *a submissive Kuwait, would fall in the hands of a dangerous dictator.*

> *Faced with such* <XNUC corr="information">*an information*</XNUC>, *the viewer is powerless.*

2.3. Nouns <GN>

2.3.1. Noun case <GNC>

<GNC> is used to tag:

- Inappropriate use of the Saxon genitive:

  > *The military* <GNC corr="recruitment system">*recruitment's system*</GNC>

  > *their* <GNC corr="student years">*student's years*</GNC>

  > *The* <GNC corr="girls'">*girls's*</GNC> *jackets*

- use of a phrase where there should be a Saxon genitive: for example, use of the 'noun + noun' or 'noun *of* noun' combinations, or in rarer cases of other prepositions, where the Saxon genitive should have been used:

  > *This is* <GNC corr="John's daughter">*John daughter*</GNC>

  > <GNC corr="My sister's car">*The car of my sister*</GNC>

  > <GNC corr="women's sexuality">*the sexuality in women*</GNC>

  > <u>Note:</u> This is an exception to the rule that we tag on the basis of the incorrect form/word only (cf. [Principle 1](#))

- omission of the genitive marker:

  > *His words, delivered in a matter-of fact tone, could almost serve as a* <GNC corr="beginner's guide">*beginners guide*</GNC> *for potential killers on "how to murder and dispose of your victims".*

<GNC corr="Hopkins's poetry">*Hopkins poetry*</GNC> *was considered too advanced...*

Warning!

In the case of *'s* or *s'* used instead of the plural morpheme *s*, the error is classified as <GNN>
(cf. Section 2.3.2.) and not <GNC>.

<GNN corr="Jeans">*Jean's*</GNN> *for sale*

2.3.2. Noun number <GNN>

<GNN> denotes errors due to the addition/omission of the plural morpheme.

Singular for plural

three <GNN corr="tenths">*tenth*</GNN> *of a mile*

People still wonder if <GNN corr="feminists">*feminist*</GNN> *have done more harm than good.*

Plural for singular

How many people are sitting in front of the television every <GNN corr="day">*days*</GNN> *at four o'clock?*

Such a <GNN corr="diagnosis">*diagnoses*</GNN> *is scary for patients.*

Errors in irregular plurals

<GNN corr="mice">*mouses*</GNN>

Use of genitive marker instead of plural -s

underprivileged <GNN corr="beings">*being's*</GNN>

Warning!

This sub-category does not include erroneously pluralised uncountable nouns, which are
tagged as lexico-grammatical errors. See Section 3.3.

<XNUC corr="evidence">*evidences*</XNUC>).

2.4. Pronouns <GP>

<GP> pertains to errors involving any of the pronoun categories: demonstrative, interrogative, personal, impersonal, possessive, reciprocal, reflexive, relative and indefinite.

2.4.1. Demonstrative pronouns <GPD>

<GPD> covers all errors involving demonstrative pronouns.

> <GPD corr="This">*That*</GPD> *is a principle of Grice about communication.*

2.4.2. Personal pronouns <GPP>

<GPP> applies to all errors involving personal pronouns. The category also includes impersonal pronouns (*it, there*).

> *Europe is a small area,* <GPP corr="it">*she*</GPP> *only represents...*
>
> <GPP corr="This">*It*</GPP> *was a short time ago and men have not had enough time to change.*

2.4.3. Possessive pronouns <GPO>

<GPO> applies to all errors involving possessive pronouns.

> *it is not easy to create a system that corresponds to the society in question.* <GPO corr="Ours">*Theirs*</GPO> *is outdated.*

2.4.4. Indefinite pronouns <GPI>

<GPI> is used to tag all errors involving indefinite pronouns (*some, any, everyone, no-one, everything, nothing,* etc.). This category also includes the pronouns *one/ones* (= substitute *one*) and the pro-form *so*.

> *I am convinced that fairness forces* <GPI corr="everybody">*anybody*</GPI> *to seek improvement.*
>
> *In Spain there is a great polemic about* <GPI corr="everything">*all*</GPI>*.*

2.4.5. Reflexive and reciprocal pronouns <GPF>

<GPF> is used to tag all errors involving reflexive and reciprocal pronouns.

> *We could hardly see* <GPF corr="each other">*one to another*</GPF> *in the fog.*

However, when imagination simply means to dream and try to relax <GPI corr="0">*themselves*</GPI>*, they reject it, preferring for example to switch on the TV.*

2.4.6. Relative and interrogative pronouns <GPR>

<GPR> pertains to all errors involving relative and interrogative pronouns.

As clearly emerges from all <GPR corr="that">*what*</GPR> *precedes.*

The government took several measures to stop the strikes, <GPR corr="which">*that*</GPR> *was not effective.*

2.4.7. Unclear pronominal reference <GPU>

Cases where the referent of the pronoun (be it a personal, demonstrative, possessive, reflexive or indefinite pronoun) is unclear are tagged <GPU>.

But there are also imprisoned people waiting for their execution who are innocent. They never had a fair trial and a real chance to get out of <GPU corr="jail">*it*</GPU>*.*

More than one lonely office worker has established a humanlike relationship with his computer. It is given a name; <GPU corr="the worker">*he*</GPU> *buys expensive "extras" to keep it entertaining when he is working overtime, staring into the deep green writing on the screen.*

Unlike the Chinese education system, the western education system demands students to assume more personal responsibility. Contrary to the Chinese <GPU corr="system">*one*</GPU>*, if you are studying at a university in Britain, …*

Warning!

Cases of missing or redundant pronouns are tagged <GP*> rather than <WM> or <WR>:

The first kind of criminal is the worst because <GPP corr="it">*0*</GPP> *is very sad that a person attempts destroys a property.*

They wish to feel <GPP corr="0">*themselves*</GPP> *smart.*

2.5. Adjectives <GADJ>

2.5.1. Adjective order <GADJO>

<GADJO> is used for all problems of adjective position. These typically involve placing the adjective after the noun it describes.

a <GADJO corr="red car">*car red*</GADJO>

The tag is also used for word-order errors when two or more adjectives precede the noun.

a <GADJO corr="small black leather">*leather black small*</GADJO> *handbag*

Some adjectives have different meanings according to whether they are used attributively or predicatively; errors arising from this are tagged as <GADJO>.

the <GADJO corr="people concerned">*concerned people*</GADJO> *are asked to fill in a form*

2.5.2. Adjective number <GADJN>

<GADJN> covers all erroneous uses of plural morphemes with adjectives.

the <GADJN corr="poor">*poors*</GADJN> *people*

The last sentences have been <GADJN corr="favourable">*favourables*</GADJN> *to women.*

less <GADJN corr="intelligent">*intelligents*</GADJN> *beings...*

2.5.3. Comparative / Superlative <GADJCS>

<GADJCS> is used for errors involving wrong formation and/or misuse of comparatives and superlatives.

Wrong formation of comparatives or superlatives of adjectives

the <GADJCS corr="richer">*more rich*</GADJCS> *our culture is ...*

the role that women should play in a <GADJCS corr="fairer">*more fair*</GADJCS> *present-day society*

Misuse of comparatives or superlatives of adjectives

He is the <GADJCS corr="youngest">*younger*</GADJCS> *of the three brothers.*

Erroneous irregular comparative/superlative adjectives

The situation is <GADJCS corr="worse">*badder*</GADJCS> *for him.*

2.6. Adverbs <GADVO>

This category covers only adverb order errors, tagged <GADVO>. Other errors involving adverbs fall within the categories of connectors <LCL>, morphology <FM>, or adverbial particles within phrasal verbs <LP>.

<GADVO> is used to tag misplaced adverbs (e.g. verb and object separated by an adverb).

> *but they* <GADVO corr="have also">*also have*</GADVO> *come to feel some changes.*
>
> *they* <GADVO corr="only see">*see only*</GADVO> *other criminals*

2.7. Verbs <GV>

2.7.1. Verb number <GVN>

<GVN> applies to all errors of agreement between a verb and its subject.

Singular subject – plural verb

> *Progress* <GVN corr="has">*have*</GVN> *been made*
>
> *Ever since the Industrial Revolution which* <GVN corr="has">*have*</GVN> *radically marked the transition from…*

Plural subject – singular verb

> *The people who employ them* <GVN corr="do not">*doesn't*</GVN> *care.*
>
> *Most of the psychological casualties* <GVN corr="happen">*happens*</GVN> *among the unprofessional soldiers.*

2.7.2. Verb Morphology <GVM>

This category pertains to erroneous uses of **verb morphology**, i.e. misuse of inflectional verbal morphemes. <GVM> includes instances such as the use of a simple past form instead of a past participle form, erroneously formed 3rd person singular morphemes or erroneous irregular past tense verbs.

> *I have* <GVM corr="swum">*swam*</GVM> *for two hours.*
>
> *I have* <GVM corr="been">*be*</GVM> *living in Paris since 1985.*
>
> *He* <GVM corr="watches">*watchs*</GVM> *television all day long.*
>
> *I have* <GVM corr="swum">*swimmed*</GVM> *for two hours.*

He <GVM corr="split">*splitted*</GVM> *his time between the two.*

It is strictly <GVM corr="forbidden">*forbaden*</GVM>*.*

2.7.3. Non-Finite/Finite verb forms <GVNF>

<GVNF> is used to tag the following errors involving non-finite/finite verb forms.

-*ing* form used instead of a relative clause

Hardly anybody from our district <GVNF corr="who is older">*being older*</GVNF> *than fifteen goes to school regularly.*

Use of a relative clause instead of -*ing* form

If you ever looked at a small someone dicing with death (…), and suddenly realized that it was your son <GNVF corr="proudly shouting">*who proudly shouts*</GNVF>*: "Mum, I'm almost there!" you would…*

You wanted to go to Glasgow but the first car <GVNF corr="that stopped">*stopping*</GVNF> *as you were hitchhiking was going to London?*

Nominal use of the bare infinitive instead of -*ing* form

By the way <GVNF corr="estimating">*estimate*</GVNF> *other cultures does not mean that you agree with everything they do.*

Nominal use of the infinitive with *to* instead of -*ing* form

For Europeans, <GVNF corr="leaving">*to leave*</GVNF> *their country is not easy.*

2.7.4. Verb voice <GVV>

<GVV> is used to tag all errors where there is confusion between the passive and the active voice.

Active for passive

It is to <GVV corr="be hoped">*hope*</GVV> *that…*

Passive for active

This seems impossible to <GVV corr="achieve">*be achieved*</GVV>*.*

These principles are worth <GVV corr="underlining">*being underlined*</GVV>*.*

2.7.5. Verb tense <GVT>

<GVT> is used to tag any misuse of tense or aspect.

Present simple for continuous

More and more companies <GVT> corr="are becoming">*become*</GVT> *multi-nationals.*

Present continuous for simple

My children <GVT> corr="quarrel">*are quarreling*</GVT> *incessantly.*

Present simple for present perfect

Recent attempts to unify Europe <GVT> corr="have been">*are*</GVT> *more peaceful.*

Simple past for present perfect

All the countries are willingly progressing towards this same aim. They <GVT> corr="have freely decided">*freely decided*</GVT> *to walk in this direction…*

Present perfect for simple past

According to an opinion poll … which <GVT> corr="was">*has been*</GVT> *published last month.*

Present simple for future

Europeans of today are so different from each other that it <GVT> corr="will still be">*is still*</GVT> *a long time before these differences disappear.*

Sequence of tenses involving the need to change more than one tense

A problem may arise with tagging a sequence of tenses – one tense change may lead to the need for other tense changes even though in the original context there is no error. In such cases, all the auxiliaries and tense forms that need changing should be tagged separately. This is an exception to Principle 6.

Last year he <GVT corr="said">*says*</GVT> *that he* <GVT corr="wanted"> *wants* </GVT> *to study medicine.*

2.7.6. Auxiliaries <GVAUX>

<GVAUX> pertains to all misuses of the primary auxiliaries (*be, have* and *do*), all modal auxiliaries except *shall* and *will*, which are classified in the tense and aspect category <GVT>, and the semi-auxiliaries (*dare, need, ought to* and *used to*).[15]

> *Three years ago, I* <GVAUX corr="was able to">*could*</GVAUX> *climb Mount Everest during the climbing championship.*
>
> *If you don't want to be the victim of a car accident and if you* <GVAUX corr="would">*0*</GVAUX> *like to enjoy your life for a few more years, please take your foot off the speed pedal.*

2.8. Word class <GWC>

<GWC> is used to tag errors due to confusion between word classes, e.g. an adjective used instead of a noun or an adverb instead of an adjective.

> *This story might have ended* <GWC corr="differently">*different*</GWC> *if silly Charles had invited Elizabeth to McDonald.*
>
> *It is* <GWC corr="ethically">*ethical*</GWC> *questionable to divide a cell a hundred times.*
>
> *Sometimes they* <GWC corr="succeed">*success*</GWC> *in escaping from their prisons.*
>
> *Not everybody is* <GWC corr="lucky">*luckily*</GWC> *enough to inherit so much money.*

3. LEXICO-GRAMMAR <X>

This category covers errors where the lexico-grammatical properties of a word have been violated. The errors result from the learner's lack of knowledge not of sentence grammar (G category) but of word grammar, i.e. the lexico-grammatical properties that are specific to a particular word (e.g. the fact that the adjective *dependent* is followed by the preposition *on* rather than *of* or that the verb *to hesitate* is followed by a *to*-infinitive rather than an *-ing* form). These properties are included under the relevant headwords in monolingual learners' dictionaries. The most important areas of error are non-finite/finite complementation of nouns, verbs, adjectives, adverbs, dependent prepositions and countable/uncountable nouns. As an <X*> error is due to insufficient knowledge of the word that triggers the error, both the triggering word and the erroneous word/pattern are tagged rather than only the triggering word/pattern. For example, *he hesitated atttending the party*

---

[15] List taken from Carter R. & McCarthy M. (2006) *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge: Cambridge University Press, p. 395.

is tagged *he* <XVCO corr="hesitated to attend">*hesitated attending*</XVCO> *the party* rather than *he hesitated* <XVCO corr="to attend">*attending*</XVCO> *the party*. This makes it possible to draw up lists of words whose patterning is particularly error-prone.

3.1. Complementation <X*CO>

This category is for complementation errors, i.e. erroneous use of the finite and non-finite patterns that complement adjectives, adverbs, nouns, prepositions and verbs.

The sub-categories are the following:

<XADJCO>: erroneous complementation of adjectives

> *I'm perfectly* <XADJCO corr="capable of going">*capable to go*</XADJCO> *on my own.*

<XNCO>: erroneous complementation of nouns

> *students have the* <XNCO corr="possibility of leaving">*possibility to leave*</XNCO>
>
> *there are* <XNCO corr="difficulties in establishing">*difficulties to establish*</XNCO> *unity*

<XPRCO>: erroneous complementation of prepositions

> <XPRCO corr="instead of going">*instead of to go*</XPRCO> *abroad*

<XVCO>: erroneous complementation of verbs

> *This will not* <XVCO corr="help people to believe">*help people believing*</XVCO> *in…*
>
> *They* <XVCO corr="enjoy gathering">*enjoy to gather*</XVCO> *in groups.*

3.2. Dependent prepositions <X*PR>

This category covers all errors involving dependent prepositions, i.e. prepositions that are intrinsically linked to a particular adjective, adverb, noun or verb. The category also includes cases of omission of a necessary preposition.

The sub-categories are the following:

<XADJPR>: adjectives used with the wrong dependent preposition

> <XADJPR corr="representative of">*representative for*</XADJPR> *their own identities*
>
> *The social sphere from which one stems appears, to a great extent, to be* <XADJPR corr="decisive in">*decisive for*</XADJPR> *one's achievements.*

\<XADVPR>: adverbs used with the wrong dependent preposition

The investigators acted \<XADVPR> corr="independently of">*independently from*\</XADVPR> *each other.*

\<XNPR>: nouns used with the wrong dependent preposition

He has a \<XNPR corr="thirst for">*thirst of*\</XNPR> *knowledge.*

The \<XNPR corr="decrease in">*decrease of*\</XNPR> *the number of violent crimes is encouraging.*

\<XVPR>: verbs used with an erroneous, missing or redundant preposition

Each European should \<XVPR corr="ask">*ask to*\</XVPR> *himself*

He is \<XVPR corr="courting">*courting to*\</XVPR> *a lady.*

They will feel the need to \<XVPR corr="move to">*move in*\</XVPR> *another country*

He is \<XVPR corr="waiting for">*waiting*\</XVPR> *Barbara.*

Warning!

This category only relates to **dependent** prepositions. In the case of errors involving **independent** prepositions (*he walked *throughout the bridge* instead of *across the bridge*), the \<LSPR> tag is used (see Section 4.1). Researchers who are interested in analysing **all** erroneous uses of prepositions should therefore consider both X*PR errors and LSPR errors.

3.3. Nouns: uncountable/countable \<XNUC>

\<XNUC> is used to tag all errors involving two specific characteristics of uncountable nouns: the impossiblity of their being determined by an indefinite article (*an information) and their lack of pluralization (*informations). Note that, in the former case, the \<XNUC> tag should be placed on both the incorrect article and the noun.

Use of the indefinite article with uncountable nouns

I need to ask you for \<XNUC corr="some information">*an information*\</XNUC>

Before embarking on \<XNUC corr="research">*a research*\</XNUC> *you will need to consider a number of factors*.

Plural of uncountable nouns

> *My parents gave me some excellent* <XNUC corr="advice">*advices*</XNUC>
>
> *Other* <XNUC corr="evidence">*evidences*</XNUC> *will be needed if we are to secure a conviction.*
>
> *The Europe of 1992 aims at* <XNUC corr="progress">*progresses*</XNUC> *in scientific cooperation.*

4. LEXIS (L)

The lexical category encompasses errors involving the conceptual and collocational properties of words or phrases. It is divided into three sub-categories: Lexical Single, Lexical Phrase and Lexical Connectors. Errors involving connectors have been placed in a separate category in view of their attested difficulty, especially in academic writing.

4.1. Lexical single <LS>

<LS> is used for conceptual (e.g. using *comprehensible* instead of *understandable*) and collocational (e.g. using *high responsibilities* instead of *heavy responsibilities*) lexical errors in **single** words. This sub-category includes solid and hyphenated compounds. Unhyphenated compounds fall under the Lexical Phrase sub-category.

Lexical errors are further subdivided according to the grammatical category of the erroneous word. The subcategories are as follows:

Lexical errors affecting adjectives <LSADJ>

> *Resorting to violence might seem like a* <LSADJ corr="understandable">*comprehensible*</LSADJ> *reaction.*
>
> *Louvain-la-Neuve is just a* <LSADJ corr="ordinary">*common* </LSADJ> *city.*
>
> *The managers took on* <LSADJ corr="heavy">*high*</LSADJ> *responsibilities.*
>
> *This is the* <LSADJ corr="main">*mean*</LSADJ> *reason, I think, why people today want a university degree.*
>
> *If society does not help them in the first stage of their criminal transformation,* <LSADJ corr="petty">*small*</LSADJ> *thieves can become professionals.*

Lexical errors affecting adverbs <LSADV>

> *Results were very different for the* <LSADV corr="fully">*wholly*</LSADV> *fit peer.*

I will try and explain it <LSADV corr="later">*onwards*</LSADV> *by providing some clear examples.*

Lexical errors affecting nouns <LSN>

Luckily, I could overcome his disappointment by giving him a <LSN corr="packet">*package*</LSN> *of stickers.*

Therefore subjects like philosophy have been eliminated from the <LSN corr="curriculum">*plannings*</LSN>*.*

Lexical errors affecting single or complex prepositions <LSPR>

My purpose <LSPR corr="in">*on*</LSPR> *this essay is to prove how we are subject to capitalism.*

Advocates of the motorway maintain that trade would prosper <LSPR corr="thanks to">*through*</LSPR> *the motorway.*

He almost died <LSPR corr="due to">*thanks to*</LSPR> *lack of oxygen.*

Warning!

Remember that the X*PR tag is used to annotate cases of erroneous **dependent** prepositions (see [Section 3.2](#)).

I'm not quite sure that I <XVPR> corr="agree with>*agree to*</XVPR> *this saying.*

Lexical errors affecting verbs <LSV>

Humanity has incessantly been put to the test but has never given up <LSV corr="asserting">*affirming*</LSV> *its infinite superiority*

Man is <LSV corr="made">*done*</LSV> *to live in society.*

You should <LSV corr="pay">*cling*</LSV> *attention to things.*

4.2. Lexical phrase <LP>

The main difference between the <LS> and <LP> tags is that <LS> applies to errors in **single words**, whereas <LP> is used when the lexical error affects **fixed word combinations**, i.e. idioms, compound units and phrasal verbs (as well as phrasal-prepositional verbs). Creating a separate category for multiword combinations is justified by the increasing importance of phraseology in language learning and teaching.

There should be a relaxation of the <LP corr="laws on genetic experimentation">*gene laws*</LP> *for science to progress.*

They adopted a <LP corr="zero tolerance">*null tolerance*</LP> *approach.*

The <LP corr="trade unions">*trading unions*</LP> *were dead against the reform.*

We must do our utmost to combat <LP corr="juvenile crime">*children's criminality*</LP>

She works as a <LP corr="sign language interpreter">*hand-speaking announcer*</LP> *at the European Union of the Deaf.*

He is a <LP corr="security guard">*guard for their security*</LP> *at the local shopping centre.*

This does not mean that thieves, bandits, muggers and so on must be <LP corr="left out">*kept away*</LP> *of these improvements of the prison system.*

In Germany one-third of all marriages <LP corr="end">*end up*</LP> *after seven years.*

If one <LP corr="looks back on">*looks back in*</LP> *former times, one realizes that there were problems.*

Warning!

1. The LP tag only comes into play if the learner uses an erroneous phrasal verb in the original text. If the learner originally used a single verb which the annotator believes should be replaced by a phrasal verb, the error is tagged <LS>, not <LP>, on the grounds that we should tag on the basis of the error, not the correction.

   It all depends on his will to <LSV corr="stand up">*stand*</LSV> *for these issues.*

   To <LSV corr="finish off">*finish*</LSV>*, I would like to emphasise that…*

2. Grammatical errors within set phrases should not be embedded in the LP tag. Instead, the whole phrase should be tagged and corrected under a single LP tag. In *How times flies*, the pluralisation error (*times* instead of *time*) should not be tagged separately. The whole phrase is considered to be erroneous, and only the LP tag is used.

   How <LP corr="time flies">*times flies*</LP>

This principle applies equally to any other type of error within a set phrase, such as the missing article in the following example.

   <LP corr="on an equal footing">*on equal footing*</LP>.

4.3. Lexis, Connectors <LC>

This sub-category consists of errors involving connectors, i.e. adverbial logical connectors <LCL>, coordinating conjunctions <LCC> and subordinating conjunctions <LCS>. The problems may be of different kinds: misuse, unnecessary use or wrong choice. Note that the tag should also be used to indicate the absence of a connector where one would be expected.

4.3.1. Logical connectors <LCL>

<LCL> covers all errors involving logical connectors, i.e. adverbs or adverbial phrases that establish a logical relationship between ideas within a sentence or between sentences. The category is further subdivided into <LCLS> for single connectors, which consist of only one word, and <LCLC> for complex connectors, which are made up of more than one word.

4.3.1.1. Single Logical Connectors <LCLS>

Many unemployed workers will find work opportunities. <LCLS> corr="Actually">*Really*</LCLS>*, I agree with this statement up to a point.*

It is not a very nice flat. <LCLS> corr="However">*Moreover*</LCLS>*, it is cheap.*

<LCLS corr="Therefore">*So*</LCLS>*, in our modern world, there is enough place for both dreams and technology.*

4.3.1.2. Complex Logical Connectors (LCLC)

<LCLC corr="To conclude">*As a conclusion*</LCLC> *it is too radical to claim that progress is good or bad for the world.*

<LCLC corr="On the contrary">*By the contrary*</LCLC> *not all improvements have been as positive as the ones previously said.*

Industrialization <LCLC corr="on the other hand">*on the other side*</LCLC> *is very dominant and powerful nowadays and controls a lot the flow of money.*

Everything what we want, we easily get, everything is on our hand. <LCLC corr="0">*By the way*</LCLC>*, science and technology exist to provide these needs and then dreams are easily achieved.*

4.3.2. Coordinating conjunctions <LCC>

<LCC> consists of all errors involving coordinating conjunctions (*and, but, or, nor, either ... or, neither ... nor*).

> *Neither Margareth* <LCC corr="nor">*or*</LCC> *John was there.*
>
> *... produce a shock (whether positive or negative) on everyone.* <LCC corr="However,">*But*</LCC> *you can only get a real idea or feeling about it if you live there.*
>
> *The latter can make convenient use of his parents' wealth by attending private courses,* <LCC corr="or">*0*</LCC> *buying useful reference tools.*
>
> <LCC corr="Neither">*Nor*</LCC> *Caesar, Napoleon nor Hitler fought to protect their own countries.*
>
> *I would not like to given up music in my life* <LCC corr="as">*and*</LCC> *it is, along with literature, a very important hobby of mine.*

4.3.3. Subordinating conjunctions <LCS>

<LCS> is used to tag all errors involving subordinating conjunctions, which include single-word conjunctions such as *after, although, as, because, before, for, if, since, that, until, when, whereas* and *while,* and complex conjunctions such as *as long as, as soon as, even if, even though, except that, in order to, in order that* and *provided that*.

> <LCS corr="Even though">*Even although*</LCS> *I did not understand a word,...*
>
> *It would be useful* <LCS corr="if">*that*</LCS> *the Forces could count on a bigger number of professional soldiers.*

Warning!

Do not confuse errors involving the relative pronoun *that* (tagged GPR) with those involving the subordinating conjunction *that* (tagged LCS).

> *If you feel emotionally affected by the poverty* <GPR corr="that">*0*</GPR> *appears on the news, you can just give some money to charity programmes.*
>
> *Even if I have not persuaded him and* <LCS corr="0">*that*</LCS> *he decides not to come and study in Louvain-la-Neuve,...*

---

5. WORD REDUNDANT, WORD MISSING, WORD ORDER (W)

5.1. Word redundant <WR>

<WR> is used to tag the inclusion of one or more unnecessary words.

> *I only want to* <WR corr="0">*to*</WR> *try and do my best.*
>
> *They are doing everything possible* <WR corr="0">*they can*</WR> *to get...*

Warning!

Note that this tag is not used for cases of erroneous insertions of articles (GA), connectors (LC*), determiners (GD*), pronouns (GP*), dependent prepositions (X*PR), and auxiliaries (GVAUX).

5.2. Word missing <WM>

This sub-category is used for errors involving omission of words. However, not all cases where a word is missing are to be labelled <WM>. Similarly to the Word Redundant category, this category does not apply to missing pronouns, determiners, dependent prepositions, articles, connectors or auxiliaries.

If there is a doubt about the missing lexical item, the item inserted as the correction can be followed by a question mark. If the missing item is not recoverable from the context, a question mark should be inserted as the correction.

> *The* <WM corr="wine">*0*</WM> *is good.*
>
> *I do not* <WM corr="see?">*0*</WM> *why I should go there.*
>
> *School has replaced family as far as education is concerned, because it is often* <WM corr="the case">*0*</WM> *that even mothers have to work full time.*

> **Warning!**
>
> The following cases are not tagged <WM>:
>
> > *He works in* <GA corr="an">*0*</GA> *office* (missing article)
> >
> > <LCLC corr="On the other hand">*On other hand*</LCLC> (missing article in a connector)
> >
> > *What we wear does not matter and neither* <GVAUX corr="does">*0*</GVAUX> *the age of the victim* (missing auxiliary)
> >
> > *Your claim that* <GPP corr="it">*0*</GPP> *is the task of the government to subsidize families is daring* (missing pronoun)

5.3. Word order <WO>

<WO> is for problems of word order that do not fall within the categories of Adverb Order <GADVO> (cf. Section 2.6.) or Adjective Order <GADJO> (cf. Section 2.5.1.).

> *The dynamic which roughly began* <WO corr="in Rome in 1957">*in 1957 in Rome*</WO>*.*
>
> *Some 30 years ago* <WO corr="the idea of … was launched">*was launched the idea of …*</WO>*.*

6. PUNCTUATION <Q>

The <Q> category includes all punctuation errors. As punctuation is a field particularly affected by personal preference, all the punctuation errors detected by the annotator will be tagged using one of the <Q*> tags (the same thing applies to connectors, cf. Section 4.3.). This means that no decisions need be made as to whether the punctuation errors are real errors or simply infelicities. This can be done by the individual analyst at a later stage if need be.

There are four subcategories of punctuation error:

<QM>: missing punctuation (e.g. missing comma)

> *John Lennon* <QM corr=",">*0*</QM> *who was born in Liverpool,…*
>
> *Fortunately* <QM corr=",">*0*</QM> *man's great capacity for adaptation…*

<QR>: redundant punctuation (e.g. redundant comma)

> *It was John* <QR corr="0">*,*</QR> *who was born in Liverpool…*
>
> *Women were seen conducting affairs and bringing negotiations to satisfactory conclusions in what men always claimed to be* <QR corr="0">*:*</QR> *"a man's world".*

<QC>: confusion of punctuation marks (comma instead of full stop, etc.); cases of comma splices are included in this category.

> *He took the books, the records, the computers* <QC corr=".">*,*</QC> *he left at eight o'clock.*
>
> *But that's a terrible lie* <QC corr=".">*,*</QC> *under no circumstance should you believe a lie like that.*
>
> *Who were the people I phoned* <QC corr="?">*:*</QC> *there was Helen, my school friend.*

<QL>: punctuation mark instead of lexical item (usually a coordinating conjunction) or lexical item instead of punctuation mark

> *He went to the seaside* <QL corr="and">*.*</QL> *He spent the whole day there.*
>
> *He took the books* <QL corr=",">*and*</QL> *the records and the computers.*

7. Infelicities <Z>

Infelicities are tagged <Z> and cover the following issues: register problems, questions of political correctness and stylistic problems. Unclear sequences of words or sentences are also tagged <Z>. The <Z> category does not include cases of clear-cut grammatical or lexical errors, but rather of words or phrases that are felt to be inappropriate, stylistically clumsy or unclear. Sometimes this can be a matter of the personal taste of the annotator.

The rationale for the creation of this category was to improve inter-rater reliability in the other categories, by excluding from them examples that are heavily reliant on the personal judgement of the annotator. Inclusion of *he/she* issues in the <GP> category may unduly increase the frequency of pronoun errors, for example, and judgements as to the acceptability or not of informal words may distort numbers in the <L> category.

The annotator should always try to analyse the problem and ensure that no other tag applies before opting for the <Z> category, in order to avoid the danger of this category becoming an ill-defined dustbin category.

1. Register problems: All register problems are tagged as infelicities <Z>. Register problems include contractions (*couldn't*), informal expressions (*a bit, lots of*), slang, interjections, use of dated vocabulary, inappropriate regionalisms, etc. Here are some examples:

> <Z corr="children">*kids*</Z> (inappropriately informal word in formal writing)
>
> *Most of them just need a job and* <Z corr="some">*a bit of*</Z> *consideration* (inappropriately informal phrase in formal writing)

this intimate place <Z corr="could not">*couldn't*</Z> *possibly be harmed by simple material conditions* (inappropriate contraction in formal writing)

There are <Z corr="two">*2*</Z> *students* (breaks the convention on spelling of numbers (1-9 expressed in words, 10 and upwards in figures)

<Z corr="pram">*perambulator*</Z> (dated language)

parity of the sexes in all walks of life – <Z corr="namely">*to wit*</Z> *at work and at home* (dated language)

<Z corr="colour">*color*</Z> (regional spelling difference – American English in a text written in British English)

<Z corr="centre">*center*</Z> (regional spelling difference – American English in a text written in British English)

2. Questions of **political correctness** include the following examples: *chair<u>man</u>* instead of *chair<u>person</u>*, *<u>man</u>kind* instead of *<u>human</u>kind*, *he/she, himself/herself,* etc.

3. The category of **stylistic problems** covers longer, foreign-sounding or clumsy chunks of text that do not contain any errors as such but that an expert writer would want to reformulate. More than in any other category, the suggested correction is likely to be one of several possible ways of rewording the passage. It will often be a matter of whole sentences or unfixed phrases, but not fixed or semi-fixed (prefabricated) phrases, which are tagged <LP>.

A person is no longer <Z corr="able to make personal decisions">*an element of personal decision*</Z>.

These two cases <Z corr="produce a vicious circle from which it is difficult to exit">*are vicious circles through which it is difficult to go*</Z>.

Also covered by stylistic problems are long-winded introductory formulations such as we can say that, it must be stated that, it must be said that:

To conclude <Z corr="0">*I shall once more insist on the fact that*</Z> *a world from which dreams and imagination are absent…*

4. The <Z> category also includes **unclear sequences of words or sentences**, where a question mark should be used instead of a correction:

Little by little, a status quo was reached and <Z corr="?">*got into a rut between both sexes*</Z>.

Therefore, by presenting the main character in such a way throughout the play, <Z corr="?">*by this close found*</Z>*, by this reversal of fortune, he makes us feel sympathy towards the character…*

Warning!

If the lack of clarity relates to pronoun reference, the tag that should be used is <GPU> (cf. Section 2.4.7.). If the intended meaning of a single word that exists in English is unclear, then the <LS> tag should be used, with a question mark as the correction (cf. Section 4.1.).

D. INDEX

LSV	**L**exical **S**ingle **V**erb
QC	Punctuation, **C**onfusion
QL	Punctuation, **L**exical
QM	Punctuation, **M**issing
QR	Punctuation, **R**edundant
WM	**W**ord **M**issing
WO	**W**ord **O**rder
WR	**W**ord **R**edundant
XADJCO	Le**X**ico-Grammar, **Adj**ective, **Co**mplementation
XADJPR	Le**X**ico-Grammar, **Adj**ective, Dependent **Pr**eposition
XADVPR	Le**X**ico-Grammar, **Adv**erb, Dependent **Pr**eposition
XNCO	Le**X**ico-Grammar, **N**oun, **Co**mplementation
XNPR	Le**X**ico-Grammar, **N**oun, Dependent **Pr**eposition
XNUC	Le**X**ico-Grammar, **N**oun, **U**ncountable / **C**ountable
XPRCO	Le**X**ico-Grammar, **Pr**eposition, **Co**mplementation
XVCO	Le**X**ico-Grammar, **V**erb, **Co**mplementation
XVPR	Le**X**ico-Grammar, **V**erb, Dependent **Pr**eposition
Z	Infelicities

E. FUZZY AREAS: PRACTICAL GUIDE

Error annotation involves a large number of fuzzy areas. The manual contains notes and warnings that alert users to these areas. In this section we have collated some of them under the following headings:

- word order
- word missing
- erroneous use of articles
- erroneous use of the plural morpheme
- erroneous use of verb morphology
- erroneous use of prepositions

WORD ORDER

1) Is it a problem of position of adjectives?
  <GADJO> cf. Section 2.5.1.
  a <GADJO corr="red car">*car red*</GADJO>

2) Is it a problem of position of adverbs?
  <GADVO> cf. Section 2.6.
  they <GADVO corr="have also">*also have*</GADVO> *come to feel some changes*

3) In all other cases:
  <WO> cf. Section 5.3.
  This roughly began <WO corr="in Rome in 1957">*in 1957 in Rome*</WO>

WORD MISSING

1) Is the missing word a connector or a word missing from a connector?
  <LC*> cf. Section 4.3.
  <LCLC corr="on the other hand">*on the hand*</LCLC>

2) Does the missing word form part of a set phrase?
  <LP> cf. Section 4.2.
  <LP corr="on an equal footing">*on equal footing*</LP>

3) Is the missing word an article?
  <GA> cf. Section 2.2.
  not only does <GA corr="the">*0*</GA> *economy take over…*

4) Is the missing word a pronoun or determiner?
  <GP*> cf. Section 2.4.
  <GD*> cf. Section 2.1.

Europe is a small area, <GPP corr="it">*0*</GPP> *only represents...*

5) Is the missing word a dependent preposition?
<X*PR> cf. Section 3.2.
We are <XVPR corr="waiting for">*waiting*</XVPR> *Barbara.*

6) In all other cases:
<WM> cf. Section 5.2.
I do not <WM corr="see?">*0*</WM> *why I should go there.*

ARTICLES

1) Is the problem due to the noun's countability status (countable/uncountable)?
<XNUC> cf. Section 3.3.
<XNUC corr="a piece of information">*an information*</XNUC>

2) Is the error contained within a connector?
<LC*> cf. Section 4.3.
<LCLC corr="on the other hand">*on other hand*</LCLC>

3) Is the error contained within a multi-word phrase?
<LP> cf. Section 4.2.
<LP corr="on an equal footing">*on equal footing*</LP>

4) In all other cases:
<GA> cf. Section 2.2.
<GA corr="a">*an*</GA> *utopia*

PLURAL MORPHEME

1) Is it an error involving the addition of a plural morpheme to an adjective?
<GADJN> cf. Section 2.5.2.
Women are seen as less <GADJN corr="intelligent">*intelligents*</GADJN> *beings.*

2) Is it a problem of regular/irregular plurals of nouns?
<GNN> cf. Section 2.3.2.
<GNN corr="mice">*mouses*</GNN>

3) Is a plural morpheme added to an uncountable noun?
<XNUC> cf. Section 3.3.
<XNUC corr="evidence">*evidences*</XNUC>

4) Is genitive *'s* or *s'* used instead of the plural morpheme?
<GNN> cf. Section 2.3.2.

<GNN corr="Jeans">*Jean's*</GNN> *for sale*

5) Is it an error involving the addition of a redundant saxon genitive to a plural noun?
    <GNC> cf. Section 2.3.1**.**
    <GNC corr="girls'">*girls's*</GNC> *for sale*

VERB MORPHOLOGY

1) Is it an error involving an erroneously formed third person singular morpheme?
    <GVM> cf. Section 2.7.2.
    He <GVM corr="watches">*watchs*</GVM> *television all day long.*

2) Is it an error involving an erroneously formed simple past morpheme?
    <GVM> cf. Section 2.7.2.
    She <GVM corr="split">*splitted*</GVM> *the biscuit and shared it.*

3) Is it an error involving the use of a simple past form instead of a past participle?
    <GVM> cf. Section 2.7.2.
    I have <GVM corr="swum">*swam*</GVM> *for two hours.*

4) Is it an error involving confusion between a bare infinitive and a past participle?
    <GVM> cf. Section 2.7.2.
    I have <GVM corr="been">*be*</GVM> *living in Paris since 1985.*

PREPOSITIONS

1) Is the erroneous preposition a dependent preposition?
    <X*PR> cf. Section 3.2.
    The latter are <XADJPR corr="dependent on">*dependent of*</XADJPR> *the former.*

2) Is the erroneous preposition an independent preposition?
    <LSPR> cf. Section 4.1.
    We can see this <LSPR corr="throughout">*along*</LSPR> *the history of humanity*

4) Is the erroneous preposition part of a connector?
    <LC*> cf. Section 4.3.
    <LCLC corr="at the same time">*in the same time*</LCLC>

5) Is the erroneous preposition part of a set phrase
    <LP> cf. Section 4.2.
    <LP corr="on equal terms">*in equal terms*</LP>