Annotating language errors in texts: investigating argumentation and decision schemas

Camille ALBERT, Laurie BUSCAIL
Marie GARNIER, Arnaud RYKNER
LLA, Université Toulouse le Mirail
31000 TOULOUSE France
mhl.garnier@gmail.com

Patrick SAINT-DIZIER
IRIT-CNRS, 118, route de Narbonne,
31062 TOULOUSE France
stdizier@irit.fr

Abstract

In this short paper, we present annotations for tagging grammatical and stylistic errors, together with attributes about the nature of the correction which are then interpreted as arguments. A decision model is introduced in order for the author to be able to decide on the best correction to make. This introduces an operational semantics for tags and related attributes.

1 Aims and Situation

Non-native English speaking authors producing documents in English often encounter lexical, grammatical and stylistic difficulties that make their texts difficult for native speakers to understand. As a result, the professionalism and the credibility of these texts is often affected. Our main aim is to develop procedures for the correction of those errors which cannot (and will not in the near future) be treated by the most advanced text processing systems such as those proposed in the Office Suite, OpenOffice and the like. In the type of errors taken into consideration, several levels are often intertwinned: morphology, lexicon, grammar, style, textual structure, domain usages, context of production, target audience, etc..

While we attempt to correct errors, it turns out that, in a large number of cases, (1) there may be ambiguities in the analysis of the nature of errors, (2) errors can receive various types and levels of corrections depending on the type of document, reader, etc., and (3) some corrections cannot be successfully done without an interaction with the author. To achieve these aims we need to produce a model of the cognitive strategies deployed by human experts (e.g. translators correcting texts, teachers) when they detect and correct errors. Our observations show that it is not a simple and straightforward strategy, but that error diagnosis and corrections are often based on a complex analytical and decisional process. Since we want our system to have a didactic capacity, in order to help writers understand their errors, we propose an analysis of error diagnosis based on argumentation theory, outlining arguments for or against a certain correction and their relative strength paired with a decision theory.

The modelling of correction strategies is based on the annotation of a large variety of types of documents in English produced by a large diversity of French speakers. Annotations allow us to identify and categorize errors as well as the parameters at stake (e.g. category change, length of new corrected segment) at stake when making corrections. This is carried out by bilingual correctors in collaboration with didacticians. Those parameters are a priori neutral in the annotation schemas. We then define a preference model that assigns polarity (positive, negative) and a weight to each of these parameters, together with additional parameters among which the target reader, the type of document, etc. An argumentation model that considers these parameters as weighted arguments, for or against a certain correction, can thus be introduced. Paired with a decision model, optimal corrections can be proposed to the author, together with explanations. This approach confers a formal interpretation to our annotation schema.

Works on the correction of grammatical errors made by human authors (Brockett, 2006), (Han et al. 2005), (Lee et al. 2006), (Tetreau et al 2008), (Writer's v. 8.2) recently started to appear. The approach presented here, which is still preliminary, is an attempt to include some didactic aspects into the correction by explaining to the user the nature of her/his errors, whether grammatical or stylistic, while weighing the pros and cons of a correction, via argumentation and decision theories (Boutiler et ali. 1999), (Amgoud et ali. 2008). Persuasion aspects also matter within the didactical perspective (e.g. Persuation Technology sym-

Proceedings of the Third Linguistic Annotation Workshop, ACL-IJCNLP 2009, pages 130–133,
Suntec, Singapore, 6-7 August 2009.

posiums), (Prakken 2006).

In this document, we present the premisses of an approach to correcting complex grammar and style errors, which allow us to evaluate difficulties, challenges, deadlocks, etc. Annotations are used here for the development of an application.

2 The annotated corpus

The documents analyzed range from spontaneous short productions, with little control and proofreading, such as personal emails or posts on forums, to highly controlled documents such as publications or professional reports. We also consider personal web pages and wiki texts. Within each of these types, we also observed variation in the control of the quality of the writing. For example, emails sent to friends are less controlled than those produced in a professional environment, and even in this latter framework, messages sent to the hierarchy or to foreign colleagues receive more attention than those sent to close colleagues. Besides the level of control, other parameters, such as style, are taken into consideration (e.g. oral vs. academic). Therefore, the different corpora we have collected form a certain continuum over several parameters (control, orality, etc.); they allow us to observe a large variety of language productions.

More details on the elaboration of corpora, definition of attributes and their stability, and annotation scenarios can be found in (Albert et al., 2009).

3 The Annotation System

Let us now briefly introduce the annotation schema we have developed. It is an ongoing effort which is gradually evaluated by real users. This schema is an attempt to reflect, in a factual and declarative way, the different parameters taken into consideration by didacticians and human translators when detecting and correcting errors. It contains several groups of tags which are given below. The values for each attribute are based on a granularity level evaluated by the didacticians of our group. They are still preliminary and require evaluation and revisions. Their structure has been designed so that they can be used in an argumentation framework.

(a) Error delimitation and characterization:
<error-zone> tags the group of words involved in the error. The zone is meant to be as minimal as possible. This tag has several attributes:
comprehension: from 0 to 4 (0 being worse): indicates if the segment is understandable, in spite of the error,
agrammaticality: from 0 to 2: indicates how ungrammtical the error is.
categ: main category of the error: lexical, syntactic, stylistic, semantic, textual,
source: calque (direct copy), overcorrection, etc.

(b) Delimitation of the correction:
<correction-zone> tags the text fragment involved in the correction. It is equal or larger than the error zone.

(c) Characterization of a given correction:
Each correction is characterized by a tag <correction> and associated attributes, positively oriented ones are underlined:
surface: size of the text segment affected by the correction: minimal, average, maximal,
grammar: indicates, whenever appropriate, if the correction proposed is the standard one as suggested by grammar rules; values are: by-default, alternative, unlikely,
meaning: indicates if the meaning has been altered: yes, somewhat, no,
var-size: is an integer that indicates the increase/decrease in number of words of the correction w.r.t. the original fragment,
change: indicates if the changes in the correction are syntactic, lexical, stylistic, semantic or textual,
comp: indicates if the proposed correction is a text fragment which is easy to understand or not; values are: yes, average, no,
fix: indicates, when mentioned, that the error is very specific to that string of words and that the correction is idiosyncratic and cannot be extended to any other such structure.
qualif: indicates the certainty level of the annotator and didacticians, it qualifies the certainty of the error detection and of the proposed correction separetely,
correct: gives the correction.

An example is the N N construction (for the sake of readability, we do not consider longer N chains), with erroneous segments like: *the meaning utterance* or *goal failure*:

It is difficult to characterize <correction-zone>
<error-zone comprehension=”2”
agrammaticality=”1”
categ=”syntax” source=”calque”>
the meaning utterance
<correction qualif=”high” grammar=”by-default”

surface= ”minimal” meaning= ”not altered” Var-size=”+2”
change=”synt” comp=”yes”
correct= ”**the meaning of the utterance**”>
</correction>
<correction qualif=”high” grammar=”unlikely”
surface= ”minimal” meaning= ”somewhat” Var-size=”0”
change=”lexical+synt” comp=”average”
correct= ”**the meaningful utterance**”>
</correction>
</error-zone> </correction-zone> **without a context**.

These tags are relatively simple and intuitive. After some basic training, 2 independent annotators covered about 25 pages (emails and reports) so that we can measure the stability of the annotations and the annotators comprehension and agreement/disagreement. Results are not easy to analyze in a simple way since annotators disagree on some error existence and nature. In about 20% of the cases we observed such forms of disagreement. Beside this observation, annotations turn out to be quite convenient, although, for each error, a considerable analysis effort is required for its analysis. Annotating texts is very much time consuming, in particular when there are several possibilities of corrections.

4 From annotations to correction rules

Our corpus (texts, emails) has been annotated following the above schema. Several steps are required in order to reach the correction rule stage of drafting rules of corrections. The approach is still exploratory, and needs further elaborations and evaluations. This is achieved through a gradual and manually controlled machine learning strategy. As a result, we get 23 main categories of errors based on the elements involved in the grammatical and stylistic aspects, e.g.: incorrect argument structure, incorrect adverb position, incorrect embedded clause construction, incorrect coordination, incorrect paragraph start.

To define a correction rule, the segment of words in the error zone first gets a morphosyntactic tagging, so that it can be easily identified as an erroneous pattern in any circumstance. All the errors that have the same erroneous pattern are grouped to form a single correction procedure. In that same category (named ’incorrect N N constructions’), another pattern is [N(+plural) N] (e.g. *horses carriage*), and it results in a different correction rule.

Concerning the pattern ’Det N N’, when all the corresponding errors are grouped, another type of correction is found that corresponds to the inversion (the predicate meaning → the meaning of the predicate). Informally, a correction rule is defined as the union of all the corrections found for that particular pattern:
(1) merge all corrections which are similar, i.e. where the position of each word in the erroneous segment is identical to the one it has in the correction; the values of the different attributes of the <correction> tag are averaged,
(2) append all corrections which have a different correction following the word to word criterion above, and also all corrections for which the attribute ’fix’ is true.
(3) tag the corrections with all the appropriate morphosyntactic details,
(4) remove the text segments or keep them as examples.

For the above example, we get the following rule:
<correction-rule>
<error-zone comprehension=”2” agrammaticality=”1”
categ=”syntax” source=”calque”
pattern=”[Det N(1) N(2)”]>
<correction qualif=”high” grammar=”by-default”
surface= ”minimal” meaning= ”not altered” Var-size=”+2”
change=”synt” comp=”yes”
web-correct= ”[Det N(1) of the N(2)]” >
</correction>
<correction qualif=”high” grammar=”unlikely”
surface= ”minimal”
meaning= ”somewhat” Var-size=”0”
change=”lexical+synt” comp=”average”
correct=”[Det Adj(deriv(N(1)) N(2)]”
exemple=”the meaningful utterance”>
</correction>
<correction qualif=”high” grammar=”by-default”
surface= ”minimal”
meaning= ”not altered” Var-size=”+2”
change=”synt” comp=”yes”
web-correct= ”[Det N(2) of the N(1)]” >
</correction> </error-zone> </correction-rule>

We observe here several competing solutions: when we have a segment like *the meaning predicate* we have no information as to the noun order and the type of preposition to insert (however, ’of’ is the most frequent one). In this example, the best solution is to use the web as a corpus. The attribute web-correct is a shortcut for a function that triggers a web search: the instanciated pattern is submitted to a search engine to evaluate its occurence frequency. The most frequent one is adopted. Other rules contain e.g. interactions with the user to get a missing argument or to correct a pronoun.

The form: *pattern → correct (or) web-correct* is a rewriting rule that operates the correction under constraints given in the 'correct' attribute and under didactic constraints given in the associated attributes. Several corrections from the same rule or from different rules may be competing. This is a very frequent situation, e.g.: the position of the adverb which may equally be either before the main verb, or at the beginning, or at the end of the sentence. A correction rule is $active$ for a given correction iff all the constraints it contains in the 'correct' attribute are met.

5 Using argumentation to structure the correction space

Our goal, within an 'active didactics' perspective, consists in identifying the best corrections and proposing them to the writer together with explanations, so that he can make the most relevant decisions. Classical decision theory must be paired with argumentation to produce explanations. In our framework, argumentation is based on the attributes associated with the tags of the correction rules. This view confers a kind of operational semantics to the tags and attributes we have defined.

Formally, a decision based on practical arguments is represented by a vector **(D, K, G, R)** defined as follows:
(1) D is a vector composed of decision variables associated with explanations: the list of the different decisions which can be taken into consideration, including no correction. The final decision is then made by the writer,
(2) K is a structure of stratified knowledge, possibly inconsistent. Stratifications encode priorities (e.g. Bratman, 1987, Amgoud et al. 2008). K includes, for example, knowledge about readers (e.g. in emails they like short messages, close to oral communication), grammatical and stylistic conventions or by-default behaviors, global constraints on texts or sentences. Each strata is associated with a weight $w_K \in [0, 1]$
(3) G is a set of goals, possibly inconsistent, that correspond to positive attributes Ai to promote in a correction. These goals depend on the type of document being written. For example, for emails, we may have the following goals: (meaning: no, comp: yes, grammar: by-default). These goals may have different weights. The form of a goal is:
$(attribute - name, value, weight)$
where weight is: $w_{Ai} \in [0, 1]$.
(4) R is a set of rejections: i.e. criteria that are not desired, e.g., for emails: (surface: not(minimal), change: style, semantic, textual). Format is the same as for G. R and G have an empty intersection. These rejections may also have weights.
Some attributes may remain neutral (e.g. var-size) for a given type of document or profile.

The global scenario for correcting an error is as follows: while checking a text, when an error pattern (or more if patterns are ambiguous) is activated, the corrections proposed in the <correction> tag are activated and a number of them become active because the corresponding 'correct' attribute is active. Then, the attributes in each of the correction, which form arguments, are integrated in the decision process. Their weight in G or R is integrated in a decision formula; these weights may be reinforced or weakened via the knowledge and preferences given in K. For each correction decision, a *meta-argument* that contains all the weighted pros and cons is produced. This meta-argument is the motivation and explanation for realizing the correction as suggested. It has no polarity.

References

Amgoud, L., Dimopoulos, Y., Moraitis, P., Making decisions through preference-based argumentation. In Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning (KR08), AAAI Press, 2008.

Bratman, M., Intentions, plans, and practical reason. Harvard University Press, Massachusetts, 1987.

Brockett et al. (2006) Correcting ESL Errors Using Phrasal SMT Techniques, Proc of COLING/ACL

Han et al., Detecting Errors in English Article Usage by Non-native Speakers, NLE, 2005

Lee, H., Seneff, A., Automatic Grammar Correction for Second-language Learners, Proc of InterSpeech, 2006

Prakken, H., Formal systems for persuasion dialogue, Knowledge Engineering Review, 21:163188, 2006.

Tetreault, M., Chodorow, C. Native Judgments of Non-native Usage, Proc of COLING Workshop on Human Judgements in Comp. Ling, 2008.