

ACCEPT

SEVENTH FRAMEWORK PROGRAMME

THEME ICT-2011.4.2(a)

Language Technologies

ACCEPT

Automated Community Content Editing PorTal

www.accept-project.eu

Starting date of the project: 1 January 2012

Overall duration of the project: 36 months

Definition of Pre-Editing Rules for English and French

Workpackage n° 2

Name: Definition of Pre- Editing and Post- Editing rules

Deliverable n° 2.2

Name: Definition of Pre-Editing Rules for English and French

Due date: 30 June 2013

Submission date: 28 June 2013

Dissemination level: PU

Organisation name of lead contractor for this deliverable: Acrolinx

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 288769.



Contents

- Objectives of the Deliverable 4
- Pre-editing Rules and Application Scenarios..... 4
- Requirements for Pre-editing Rules in ACCEPT 4
- Evaluation Methodology 5
- English 5
 - Description of English Rules (Symantec and TSF)..... 5
 - English Evaluation Methodology 6
 - English Rules for Translation into French 6
 - English Rules for Translation into German 7
 - Results for the English Rules 7
 - Forum Posts: English to German 7
 - Forum Posts: English to French 7
 - TSF Documents: English to French 8
- French..... 9
 - Description of French Rules..... 9
 - Symantec Rules 9
 - TSF Rules..... 10
 - Evaluation Methodology 11
 - Results for the French Rules..... 11
 - Symantec 11
 - TSF 12
- Conclusions..... 13
- References..... 14
- Appendixes I and II 15

Definition of Pre-Editing Rules for English and French

Objectives of the Deliverable

The first half of WP2 of the ACCEPT project covers the development of pre-editing rules for both English and French. This deliverable describes the different methods we have used for the development of pre-editing rules, the evaluation methodology, and gives the results in the form of the finally selected rules and their impact both in English and in French.

Pre-editing Rules and Application Scenarios

A pre-editing rule is a textual pattern that reformulates the source text in order to improve the translatability and output of the statistical machine translation system (SMT) being developed in WP4. In ACCEPT these rules are implemented in the Acrolinx rule formalism. More details on the syntax of Acrolinx rules can be found in deliverable D 2.1.

Between M1-M18, we have developed pre-editing rules for the five translation scenarios covered by ACCEPT, shown in Table 1. For Symantec, we focused on Norton forums and for TSF on field reports from MDM (*Médecins du Monde*) and MSF (*Médecins sans Frontières*).

Content type	Source language	Target language
Norton forum posts	English	French
Norton forum posts	English	German
Norton forum posts	French	English
TSF documents (MDM), field reports	English	French
TSF documents (MDM/MSF), field reports	French	English

Table 1. Pre-editing scenarios in ACCEPT

Requirements for Pre-editing Rules in ACCEPT

Acrolinx already comes with a large set of rules designed to improve the quality of a given text in terms of spelling, grammar, style, terminology, and overall readability. Typically, these rules are meant to be used in an interactive workflow, for example by providing users with a list of correction options to choose from, or explanations on what to do in case no correction can be generated.

In contrast, the ACCEPT context imposes substantially different requirements on the rules, described in detail in deliverable D 2.1:

- The primary goal is to improve the quality of the MT output, which only partially overlaps with the goal of improving the input text quality.
- Depending on the community, some users may be less willing to interact manually with pre-editing rules than others, implying that automatic rules are preferable.

These observations led us to develop three types of pre-editing rules that will be described below in detail for each language:

1. Rules that improve the input text and always provide exactly one suggestion. These rules can be applied automatically. The improved text can be shown to users.
2. Rules that improve the input text further, but give no or many suggestions. Therefore, they require interaction with the user who creates content.
3. Rules that negatively affect the input text while improving the MT output. These rules always have a unique suggestion, which is automatically applied. The output of such rules is passed directly to the MT system, and never shown to users.

Evaluation Methodology

We have employed a similar manual comparative evaluation methodology for both English and French, using the Amazon Mechanical Turk (AMT) platform (see first ACCEPT annual report). The aim of the task was to evaluate rules individually by comparing the translations of pre-edited and non-pre-edited sentences; the judge is presented with the source sentence and both translations, and asked to give a preference on a three- or five-point scale. The details, which differ slightly for the various scenarios, are explained later.

This manual methodology was first motivated by the fact that we had no translation references. In further work (Gerlach et al., 2013 and Gerlach et al., submitted), we showed that AMT judgments correlate with translator judgments and that AMT judgments on improvement of translation quality as a result of pre-editing correlate with strong decreases in post-editing time, which is what we want to reach in the ACCEPT project. These results demonstrate that the methodology is appropriate for measuring the success of this WP.

The impact of all the rules together will be evaluated in WP9 following the same methodology for all language pairs, in *Task 9.1*.

English

Description of English Rules (Symantec and TSF)

The first goal in English was to evaluate to what extent the pre-existing rules could be re-used for SMT systems. In M1-12, we have evaluated the rules using a fully automated framework that uses automatic metrics to measure the impact of pre-editing rules with respect to given reference translations (see Roturier et al., 2012).

In general, we found that only a few existing rules have a predictably positive or negative effect on the translation quality. The most effective rules perform local changes: correcting spelling issues, terminology usage, punctuation problems, missing spaces, etc. More general linguistic improvements of the source text, however, often have a seemingly random impact on the MT output. Nevertheless, we were able to identify a number of Acrolinx grammar and style rules that improve the MT output. Table 2 shows examples for the impact of these rules on translation to German and French.

	Raw	Pre-edited
Source	How much longer until <i>its</i> fixed?	How much longer until <i>it's</i> fixed?
SMT output	Wie viel länger, bis <i>seine</i> behoben?	Wie viel länger, bis <i>es</i> behoben?
Source	It is possible to <i>setup</i> Custom Power Schemes to use.	It is possible to <i>set up</i> Custom Power Schemes to use.
SMT output	Il est possible de <i>le programme d'installation personnalisée</i> les différents modes de gestion de l'alimentation à utiliser.	Il est possible de <i>configurer</i> les différents modes de gestion de l'alimentation <i>personnalisé</i> à utiliser.

Table 2. Examples for rules that improve the source and the translation

Following the observation that local changes have the most deterministic impact, we have thus developed between M12-16 a set of lexical reformulations that are highly specialized for the ACCEPT task (Symantec and TSF). This includes splitting sentences when specific word patterns occur, or removing or changing certain single words. We have implemented these reformulations in a single Acrolinx rule called “reformulation”. These reformulations (or rule suggestions) do not necessarily improve the input text. Table 3 shows examples for such reformulations. A complete list of all the TSF and Symantec English rules can be found in Appendix 2.

	Raw	Pre-edited
Source	You <i>have to</i> remember that Norton is not perfect.	You <i>must</i> remember that Norton is not perfect.
SMT output	Sie <i>haben, daran zu erinnern,</i> dass Norton ist nicht perfekt.	Sie <i>dürfen nicht vergessen,</i> dass Norton ist nicht perfekt.
Source	<i>So</i> it's not always the obvious cause.	<i>Therefore</i> it's not always the obvious cause.
SMT output	Il n'est pas toujours évident la cause.	<i>Par conséquent,</i> il n'est pas toujours évident la cause.

Table 3. Examples for rules that improve the translation only

English Evaluation Methodology

English Rules for Translation into French

We extracted 10,000 sentences from forum data and about 6,000 sentences from the TSF documents. For this evaluation, we only looked at the new automatic rules developed for this project: we ran Acrolinx on these sentences and extracted the flags with exactly one suggestion. We restricted the extraction to at most 50 flags per rule. We then automatically created the reformulated (or “pre-edited”) sentence for each original (or “raw”) sentence. The sentence pairs were then translated using the English-French Moses translation baseline system developed in WP4 (deliverable D 4.1). This produced pairs of the translated raw and the translated pre-edited sentences. Sentences where the translations were identical were marked as “no impact”.

As described above, the comparative evaluation was performed using the Amazon Mechanical Turk (AMT) platform. Evaluators were shown the raw (non-pre-edited) English source sentence and the two French translations, and asked to rate the two translations with respect to the English original phrase on a five-point scale: “first clearly better”, “first slightly better”, “about the same”, “second slightly better”, “second clearly better”. To prevent bias, the order of the raw and the pre-edited translations was swapped randomly.

Each sentence was rated independently by three different evaluators. Majority voting was used to reduce the results to a three-point scale: “raw better”, “same”, and “pre-edited better”. If no majority judgment could be reached, the sentence was marked as “no majority judgment” instead.

English Rules for Translation into German

For this evaluation, the same English sentence pairs as for the English-French forum evaluation were used. The English sentences were translated using the English-German baseline system, and evaluated by a student. The raw English sentence and the German translations of the raw and the pre-edited sentence were presented to the student, who was asked to rate them on a three-point scale (“first better”, “same”, “second better”). The rating “same” also included instances where the translations were identical. As in the other evaluations, the translations were swapped randomly. Unrated sentences were marked as “unclear”.

Results for the English Rules

Forum Posts: English to German

The results of the evaluation are shown in Table 4. Automatic rules flagged around 18% of the sentences (1831 flags for 10,000 sentences). For the evaluated set (342 sentences), the pre-editing rules improve the MT quality in 49% of the sentences, whereas they have a negative effect only in 24%, which is statistically very significant ($p < 0.0001$). The sample size was however not large enough to be able to draw conclusions about the efficacy of individual rules, as was done for the French to English rules reported later; the only important exception was that the spelling rules with unique suggestions did produce a significant improvement ($p < 0.05$).

Rule name	Flags/10k sentences	Evaluated flags	Raw better	Same + no impact	Pre-edited better	No majority judgments
reformulation	1049	50	30%	16%	54%	0%
spelling (unique suggestions)	516	50	22%	14%	60%	4%
noun_adjective_confusion	74	50	18%	56%	22%	4%
use_end_of_sentence_punctuation	47	47	28%	11%	60%	2%
use_will	35	35	29%	26%	43%	3%
its_it_is_confusion	34	34	29%	12%	59%	0%
a_an_distinction	27	27	15%	52%	26%	7%
use_comma_after_introduutory_phrase	20	20	25%	30%	40%	5%
avoid_duplicates	11	11	9%	27%	64%	0%
than_then_confusion	8	8	25%	13%	63%	0%
use_verb_with_ing_verb	5	5	20%	20%	60%	0%
incorrect_preposition	3	3	0%	0%	100%	0%
use_comma_with_parenthetical_expressions	2	2	0%	0%	100%	0%
Total	1831	342	24%	25%	49%	3%

Table 4. Evaluation of English pre-editing rules for translation of forum texts to German

Forum Posts: English to French

When translating the same sentences to French, a very different set of rules proved to be effective. Table 5 shows the results. In total, the rule flagged 24% of sentences. For the evaluated set, they improved the translations of 40% of the pre-edited sentences, while they have a negative effect in

28% of the cases, which is statistically significant ($p < 0.01$). It is notable that the spelling corrections tend to have a negative effect on the French translations, while they had a positive effect on the German translations of the same sentences. Investigating this is subject of future work.

TSF Documents: English to French

For this scenario, 45% of flags improve the MT output, whereas 20% have a negative effect (see Table 6): this is, again, very significant ($p < 0.0001$) Even considering the fact that we only evaluated about 6000 sentences and did not include the spelling flags, the total flag count is notably lower than in the forum scenario, reflecting the overall higher source quality (352 flags on 6000 sentences, 6%).

Rule name	Flags/10k sentences	Evaluated flags	No impact	Raw better	Same	Pre-edited better	No majority judgements
reformulation	1049	50	24%	34%	0%	36%	6%
capitalize_at_beginning_of_sentence	619	50	52%	14%	8%	22%	4%
spelling (unique suggestions)	516	50	18%	34%	6%	28%	14%
subject_verb_agreement	57	50	12%	32%	2%	44%	10%
use_end_of_sentence_punctuation	47	47	0%	36%	2%	51%	11%
noun_adjective_verb_confusion	42	42	31%	21%	2%	33%	12%
avoid_colloquialism_and_metaphorical_language	35	35	6%	23%	6%	54%	11%
its_it_is_confusion	34	34	9%	32%	0%	38%	21%
np_number_agreement	25	25	4%	32%	4%	52%	8%
avoid_duplicates	11	11	18%	18%	0%	46%	18%
wrong_word	9	9	11%	33%	0%	44%	11%
amount_number_confusion	2	2	0%	0%	0%	100%	0%
missing_word	2	2	0%	0%	0%	100%	0%
to_too_confusion	2	2	0%	0%	0%	50%	50%
use_verb_with_to_and_infinitive	1	1	0%	0%	0%	100%	0%
adjective_adverb_confusion	1	1	0%	0%	0%	100%	0%
Total	2452	411	18%	28%	3%	40%	11%

Table 5. Evaluation of English pre-editing rules for translation of forum texts to French

Rule name	Flags/6k sentences	Flags evaluated	No impact	Raw better	Same	Pre-edited better	Unclear
capitalize_at_beginning_of_sentence	110	50	46%	14%	2%	34%	4%
noun_adjective_verb_confusion	50	50	32%	12%	0%	46%	10%
use_end_of_sentence_punctuation	49	49	0%	31%	0%	55%	14%
use_comma_after_introduductory_phrase	36	36	50%	8%	0%	36%	6%
reformulation_tw	29	29	7%	34%	3%	38%	14%
np_number_agreement	25	25	16%	32%	4%	44%	4%
use_comma_after_subordinate_phrase	24	24	38%	17%	0%	29%	17%
do_not_use_whether_or_not	9	9	0%	11%	0%	78%	11%
amount_number_confusion	6	6	0%	33%	0%	50%	17%
where_were_confusion	5	5	0%	20%	0%	80%	0%
wrong_word	3	3	0%	0%	0%	100%	0%
avoid_duplicates	2	2	0%	0%	0%	100%	0%
its_it_is_confusion	2	2	0%	0%	50%	50%	0%
use_verb_with_to_and_infinitive	2	2	50%	0%	0%	50%	0%
Total	352	292	25%	20%	1%	45%	10%

Table 6. Evaluation of English pre-editing rules for translation of TSF documents to French

French

Description of French Rules

Symantec Rules

Between M1 and M12, a stable set of rules with significant positive impact was developed from scratch for Symantec data (Gerlach et al., 2013). The rules focus mainly on four phenomena, which proved to be troublesome for SMT of French forums into English: word confusion (due to homophones), informal and familiar French, punctuation, and local structural divergences between French and English. The main criteria for their definition have been precision and impact on translation into English.

The rules are grouped into **three sets**. Some of the rules treat unambiguous cases and have unique suggestions. These are therefore grouped in a set (**Set 1**) which can be applied automatically with no human intervention. This contains rules for homophones, word confusion, tense/mode confusion, elision and punctuation (in particular, hyphenation). Examples are shown in Table 7.

	Raw	Pre-edited
Source	oups j'ai oublié , j'ai sa aussi.	oups j'ai oublié , j'ai ça aussi.
SMT output	Oops I forgot, I have its also.	I have forgotten, I have this too.
Source	avez vous des explications ou astuces pour que cela fonctionne?	Avez-vous des explications ou astuces pour que cela fonctionne?

SMT output	Have you explanations or tips for it to work?	Do you have any explanations or tips for it to work?
------------	---	--

Table 7. Examples for rule Set 1

The remaining of the rules for humans have either multiple suggestions or no suggestions, thus requiring human intervention. These are grouped in a second set (**Set 2**), which contains spelling and rules for agreement (subject-verb, noun phrase and past participle agreement) and style, mainly for correcting informal/familiar language (for example, for avoiding cleft sentences, direct questions, use of present participle, incomplete negation, abbreviations). An example is shown in Table 8.

	Raw	Pre-edited
Source	Tu as lu le tuto sur le forum?	As-tu lu le tutoriel sur le forum?
SMT output	You have read the Tuto on the forum?	Have you read the tutorial on the forum?

Table 8. Example for rule Set 2

Finally, a third set (**Set 3**) contains the rules for the machine that should not be visible to end-users. The rules in this set modify word order and badly translated words or expressions to produce variants better suited to MT. One important rule converts the informal second person (*Tu as compilé ?*) into its formal correspondent (*Vous avez compilé ?*), which is much more frequent in the training data (Rayner et al., 2012). Another rule deals with French clitics (*le, nous, vous, me*), moving them after the verb as in English and reformulating. Examples are shown in Table 9.

	Raw	Pre-edited
Source	J'ai apporté une modification dans le titre de ton sujet.	J'ai apporté une modification dans le titre de votre sujet
SMT output	I have made a change in the title of tone subject	I have made a change in the title of your issue
Source	Il est recommandé de la tester sur une machine dédiée.	Il est recommandé de tester ça sur une machine dédiée.
SMT output	It is recommended to the test on a dedicated machine.	It is recommended to test it on a dedicated machine.

Table 9. Example for rule Set 3

TSF Rules

From M15 to M18, the Symantec rules were adapted for TSF data (field reports from MDM and MSF). Since pre-edition at TSF will be done by linguists or translators and since texts contain fewer mistakes and are less convergent, all the rules for humans were merged into one set (**Set 1**), intended to be applied manually by users. **Set 2** includes all the automatic rules for machine.

Both manual and automatic Symantec rules were first specialized by adding specific abbreviations, anglicisms, terminology, etc. Some new rules were then added to deal with specificities of the NGO's content. Manual rules (**Set 1**) were developed to simplify long and difficult sentences, very frequent in the TSF data, for example moving the subject before the verb or cutting the sentence. One of the main translation problems is related to the position of adjectives, which differs in English and French. Several automatic rules in **Set 2** move the adjective(s) before the noun. Another rule in this set replaces the simple past (*commencèrent*) by the imperfect (*commençaient*), more frequent in the training data. Examples of the application of the two sets are shown in Table 10 and Table 11.

	Raw	Pre-edited
Source	De la même façon, la traduction birmane de " Referral Fund " sous entend que son utilisation est limité aux urgences médicales.	De la même façon, la traduction birmane de " Referral Fund " sous-entend que son utilisation est limité aux urgences médicales.
SMT output	In the same way, the Burmese translation of 'Referral Fund' under meant that its use is limited to medical emergencies.	In the same way, the Burmese translation of 'Referral Fund' implies that its use is limited to medical emergencies.
Source	Ambulances avec chauffeurs et infirmiers disponibles 24h/24, reliés à la flotte téléphonique existant entre toutes les structures	Ambulances avec chauffeurs et infirmiers disponibles 24h/24, reliés à la flotte téléphonique qui existe entre toutes les structures
SMT output	Ambulances with drivers and nurses available 24h / 24, linked to the fleet between all <i>the</i> structures existing telephone	Ambulances with drivers and nurses available 24h / 24, linked to the <i>telephone</i> fleet that exists between all structures

Table 10. Examples for rule Set 1

	Raw	Pre-edited
Source	Renforcement des capacités du personnel MSHP à travers des formations	Renforcement des capacités du MSHP personnel à travers des formations
SMT output	Strengthen the capacity of the staff MSHP through training	Strengthen the capacity of the MSHP staff through training
Source	Les médecins nigériens se sentirent trahis.	Les médecins nigériens se sentaient trahis.
SMT output	The Nigerian doctors are sentirent betrayed.	The Nigerian doctors felt betrayed.
Source	Cela a permis d'augmenter de façon exponentielle le nombre d'enfants soignés.	Cela a rendu possible d'augmenter de façon exponentielle le nombre d'enfants soignés.
SMT output	This helped to increase exponentially the number of children treated.	This has made it possible to increase exponentially the number of children treated.

Table 11. Examples for rule Set 2

Evaluation Methodology

In French, we extracted 10000 sentences from forum data and 2500 sentences from the TSF documents in the selected subdomain. We have used a similar evaluation methodology with AMT judges as in English, with three main differences: 1) the evaluation included both manual and automatic rules without spelling, 2) the rule sets were applied in sequence, i.e. each set took as input the result of the previous set (for Symantec, Set 1 and 3 were applied automatically, while Set 2 was applied manually by a student; for TSF, Set 1 was applied manually before the automatic Set 2), and 3) evaluators were given the pre-edited source, excepted for automatic rules which degrade the source (**Set 3** for Symantec and **Set 2** for TSF).

Results for the French Rules

Syantec

Table 12 and Table 13 summarise the results for Symantec by set and by category of rules. Complete and up-to-date results per each rule can be found in Appendix I.

Table 12 shows that the three sets of rules produce a significant impact on translation for the subset of evaluated sentences. **Set 1** flagged 42% of the 10,000 sentences. Even if it was applied automatically, it improved the quality of translation in 49% of the evaluated sentences and had a negative impact in only 10%. This can be interpreted roughly as the noise produced by this set (false alarms), which amounts to 9%. **Set 2** flagged 20% of the cases. For the evaluated sentences of this set, the impact was positive in 58% of cases and negative in 17%. As expected, the precision for this

manual set is lower (88%). Finally, **Set 3** flagged 36% of cases and improved the quality in 58% of the cases. The precision of this automatic set is very high (98%).

	precision	Flags 10K sentences on sentences	Evaluated flags	No impact	Raw better	About the same	Preedited better	No majority judgement	impact	significant p<0.05
Symantec (10,000 sentences)										
Set 1	91%	42%	611	30%	10%	8%	49%	4%	pos	yes
Set 2	88%	20%	674	15%	17%	6%	58%	4%	pos	yes
Set 3	98%	36%	448	19%	17%	6%	53%	5%	pos	yes

Table 12. Evaluation by set of rules

The Table 13 presents the results on the 10,000 sentences grouped by category of rules and sorted by frequency. Categories include for **Set 1** and **Set 2**, 1) **punctuation**, including the rule for elision, 2) **grammar (agreement)**, 3) **grammar (tense)**, with rules for correcting tenses/modes, 4) **homophones**, with all the rules for confusion at the level of syntactic categories, 5) **informal** style. For **Set 3**, we have: 1) **clitics**, with all the rules for reordering clitics, 2) **reformulation**, which replaces an expression by another, 3) **word order** for other phenomena than clitics.

All categories produce a significant impact on the baseline except the category **grammar (tense)**. This one has a negative impact since it tends to produce tenses/moods not well covered by the training data (subjunctive, conditional, etc.). Among the significant rules, the most frequent concern punctuation, informal second person singular, informal language and homophones; the less frequent concern word order, reformulation and grammar (agreement).

Rule category	Flagged sentences	Included in evaluation	No impact	Raw better	About the same	Preedited better	No majority judgement	Impact	Significant	p-value
punctuation	3796	416	35%	8%	8%	44%	5%	pos	Yes	$p < 0.0001$
"Tu"	1968	50	40%	6%	8%	42%	4%	pos	Yes	$p < 0.00015$
clitics	1206	150	21%	18%	9%	46%	5%	pos	Yes	$p < 0.001$
informal	971	367	11%	18%	5%	59%	6%	pos	Yes	$p < 0.0001$
homophones	659	323	17%	12%	10%	57%	4%	pos	Yes	$p < 0.0001$
grammar (agreement)	591	150	21%	15%	7%	55%	2%	pos	Yes	$p < 0.0001$
reformulation	177	177	11%	15%	5%	65%	5%	pos	Yes	$p < 0.0001$
word order	71	71	17%	24%	4%	48%	7%	pos	Yes	$p < 0.01$
grammar (tense)	36	28	32%	32%	7%	25%	4%	pos	No	0.42

Table 13. Evaluation by rule category

TSF

Table 14 and Table 15 summarize the results for TSF by set and by category of rules. Complete and up-to-date results per each rule can be found in Appendix I.

In general, the results are very similar to those obtained for French Symantec data, even if the rules are different. Table 14 shows that the two sets of rules for TSF also produce a significant impact on translation for the subset of evaluated sentences. **Set 1** flagged 36% of the 2,500 sentences. It improved the quality of translation in 49% of the evaluated sentences and had a negative impact in

only 13% of the cases. **Set 2** flagged 62% of sentences. For the evaluated sentences of this set, the impact was positive in 54% of cases and negative in 20% of cases.

	precision	Flags/2,5K sentences	Evaluated flags	No imoacr	Raw better	About the same	Preedited better	No imajority judgement	impact	significant p<0.05
TSF (2,500 sentences)										
Set 1	88%	36%	486	26%	13%	3%	49%	10%	pos	yes
Set 2	99%	62%	392	19%	20%	3%	54%	5%	pos	yes

Table 14. Evaluation by set of rules

Table 15 gives the results by category. All categories have a significant impact, except those dealing with informal language, not so frequent in this domain, and with agreement (most of the flags concern NP agreement, which seems to have no impact on SMT). The most frequent and significant rule in this domain is related to word order (which includes the rules for changing the position of adjectives). Two rule categories were added for the TSF domain: **case** and **simplification**, which contain the rules for simplifying long sentences, both highly significant. Rules for clitics and homophones are significant, but clearly less frequent than for Symantec.

Rule category	Flagged sentences	Included in evaluation	No impact	Raw better	About the same	Preedited better	No majority judgement	Impact	Significant	p-value
word order	1370	219	19%	21%	1%	54%	5%	pos	yes	$p < 0.0001$
punctuation	288	204	35%	12%	4%	38%	10%	pos	yes	$p < 0.0001$
case	265	100	19%	10%	6%	63%	2%	pos	yes	$p < 0.0001$
simplification	167	84	7%	11%	1%	76%	5%	pos	yes	$p < 0.0001$
grammar (agreement)	111	57	32%	30%	4%	30%	5%	-	no-	-
reformulation	91	91	21%	20%	3%	53%	3%	pos	yes	$p < 0.001$
homophones	74	58	16%	10%	3%	57%	12%	pos	yes	$p < 0.0001$
clitics	49	49	16%	22%	2%	53%	6%	pos	yes	$p < 0.05$
Informal (+ "Tu')	18	17	35%	12%	6%	35%	12%	pos	no	0.28

Table 15. Evaluation by rule category

Conclusions

We draw the following main conclusions from the above experiments:

1. Pre-editing rules can have a substantial effect on translation quality. There is a good correlation between manual contrastive evaluation and improvement in post-editing times.
2. Effective pre-editing rules are usually local (in many cases lexical), and need to be made domain-specific. This is clearly shown by the contrasting results on the TSF data. For French-English, the Symantec rules were manually adapted to TSF, requiring about two person-weeks of effort; they performed nearly as well on the new domain. For English-French, the Symantec rules were used in almost unadapted form, and results were much weaker.

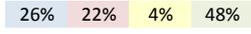
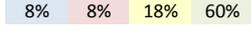
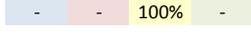
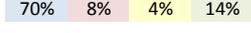
3. Comparisons with other methods, included in WP4, suggest that pre-editing is competitive with data-driven methods. In the two studies carried out so far, pre-editing was either slightly better than these methods (Rayner et al., 2012) or about the same (Bouillon et al., 2013).

References

- Pierrette Bouillon, Johanna Gerlach, Ulrich German, Barry Haddow, and Manny Rayner. Two Approaches to Correcting Homophone Confusions in a Hybrid Machine Translation System.
In *Proceedings of Second Workshop on Hybrid Approaches to Translation (HyTra)*, Nice, France, August 2013.
- Johanna Gerlach, Victoria Porro, Pierrette Bouillon and Sabine Lehmann. La prédiction avec des règles peu coûteuses, utile pour la TA statistique des forums ?
In *Proceedings of 20ème conférence du Traitement Automatique du Langage Naturel (TALN)*, Sables d'Olonne, France, June 2013.
- Johanna Gerlach, Victoria Porro, Pierrette Bouillon and Sabine Lehmann (submitted). Combining pre-editing and post-editing to improve SMT of user-generated content.
- Manny Rayner, Pierrette Bouillon and Barry Haddow. Using Source-Language Transformations to Address Register Mismatches in SMT.
In *Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA)*, San Diego, October 2012.
- Johann Roturier, Linda Mitchell, Robert Grabowski and Melanie Siegel. Using Automatic Machine Translation Metrics to Analyze the Impact of Source Reformulations.
In *Proceedings of the Conference of the Association for Machine Translation in the Americas (AMTA)*, San Diego, October 2012.

Appendix I: French pre-editing rules. Rule sets, examples, evaluation results

Symantec Set 1 (for humans, unique suggestions, automatic application)

	Flags	Rule Precision	Rule Impact	
	Total flags on 10k sentences % of rule application on 10k sentences	Evaluated Flags Wrong Flags Precision	Evaluated Flags No impact on translation Negative impact About the same Positive impact Unclear (no majority) Rule impact	Statistically significant p-value <0.05
a_vs_a	Confusion of the homophones "a", "as" and "à" Avez-vous réussi a faire l'installation de NIS 2010 en français? Avez-vous réussi à faire l'installation de NIS 2010 en français? 349 3.49%	50 - 100%	50 	pos yes
ajoutez_un_blanc	Missing space J'ai essayé d'uploader un fichier de 4ko seulement, et le problème est le même. J'ai essayé d'uploader un fichier de 4 ko seulement, et le problème est le même. 109 1.09%	50 3 94%	50 	pos no
ajoutez_un_trait_d_union	Missing hyphenation c'est à dire ? c'est - à - dire ? 186 1.86%	50 - 100%	50 	pos yes
ajoutez_une_virgule	Missing comma je ne comprend pas ce qui s'est passé mais j'ai trois comptes Norton. je ne comprend pas ce qui s'est passé, 372 3.72%	50 4 92%	50 	pos yes
ca_vs_ca	Confusion of the homophones "ça" and "sa" Merci beaucoup, je fais sa de suite Merci beaucoup, je fais ça de suite 16 0.16%	16 1 94%	16 	pos yes
ce_vs_se	Confusion of the homophones "ce" and "se" C'est se qui s'est passé il y un an lorsque j'ai changé mon ordinateur. C'est ce qui s'est passé il y un an lorsque j'ai changé mon ordinateur. 43 0.43%	43 8 81%	43 	pos yes
ci_vs_si	Confusion of the homophones "ci", "si" and "s'y" Cela permet, même si on si prend à la dernière minute, de recevoir les mises à jour. Cela permet, même si on s'y prend à la dernière minute, de recevoir les mises à jour. 1 0.01%	1 - 100%	1 	0
confusion_futur_conditionnel	Use of future tense instead of conditional Je souhaiterai savoir s'il existe un moyen de récupérer les paramètres de sauvegarde ? Je souhaiterais savoir s'il existe un moyen de récupérer les paramètres de sauvegarde? 2 0.02%	2 - 100%	2 	pos no
des_vs_des	Confusion of the homophones "des", "dès" and "dés" Dés que je le peux, je fais ce que tu me conseilles. Dès que je le peux, je fais ce que tu me conseilles. 4 0.04%	4 - 100%	4 	pos no
du_vs_du	Confusion of the homophones "du" and "dû" Tu as du mettre une protection par mot de passe, ça ne se met pas par défaut. Tu as dû mettre une protection par mot de passe, ça ne se met pas par défaut. 31 0.31%	31 1 97%	31 	pos yes
elidez_ce_mot	Missing or incorrect elision Est-ce que il s'agit ici de Norton Online Backup ou du backup de Norton 360? Est-ce qu'il s'agit ici de Norton Online Backup ou du backup de Norton 360? 129 1.29%	50 1 98%	50 	pos no

erreur_de_majuscule	Wrong capitalisation <i>Norton Ghost est plus complet et Te permet de faire un image sur un disque dur ou sur un DVD.</i> <i>Norton Ghost est plus complet et te permet de faire un image sur un disque dur ou sur un DVD.</i>	1 0.01% 1 - 100%	1 100% - - - - 0 0
espaces_autour_ponctuation	Incorrect spaces before or after punctuation <i>Comme l'a stipulé Misstigrÿ , il s'agit de services.</i> <i>Comme l'a stipulé Misstigrÿ , il s'agit de services.</i>	2430 24.30% 50 - 100%	not evaluated
evitez_conditionnel	Incorrect use of conditional tense <i>Il serait très utile si vous pourriez temporairement restaurer les serveurs DNS d'Orange.</i> <i>Il serait très utile si vous pouviez temporairement restaurer les serveurs DNS d'Orange.</i>	3 0.03% 3 2 33%	3 33% 67% - - - neg no
evitez_ponctuation	Unnecessary punctuation <i>Reviens nous voir si tu as d'autres questions ,et tiens nous au courant du résultat !</i> <i>Reviens nous voir si tu as d'autres questions et tiens nous au courant du résultat !</i>	258 2.58% 50 7 86%	50 4% 16% 18% 48% 14% pos yes
expression_incorrecte	Spelling error in common expression <i>La licence de ton produit est quand à elle sauvegardée sur ton compte Norton en ligne.</i> <i>La licence de ton produit est quant à elle sauvegardée sur ton compte Norton en ligne.</i>	17 0.17% 17 5 71%	17 35% 18% 6% 29% 12% pos no
la_vs_la	Confusion of the homophones "la" and "là" <i>Par contre, la je ne vois plus vraiment de solution pour ce souci ...</i> <i>Par contre, là je ne vois plus vraiment de solution pour ce souci ...</i>	54 0.54% 50 14 72%	50 8% 20% 20% 44% 8% pos no
ma_vs_m_a	Confusion of the homophones "ma" and "m'a" <i>j'ai fait comme tu ma dit plus haut j'ai enlevé Norton 360 et j'ai réinstallé Norton 360.</i> <i>j'ai fait comme tu m'as dit plus haut j'ai enlevé Norton 360 et j'ai réinstallé Norton 360.</i>	1 0.01% 1 - 100%	1 - - - 100% - pos no
mettez_imperatif	Use of indicative instead of imperative (homophones) <i>CD inséré, quand tu démarres le PC, regardes en bas si tu vois "Boot..."</i> <i>CD inséré, quand tu démarres le PC, regarde en bas si tu vois "Boot..."</i>	2 0.02% 2 - 100%	2 - - - 100% - pos no
ou_vs_ou	Confusion of the homophones "ou" and "où" <i>d'ou ma question...</i> <i>d'où ma question...</i>	8 0.08% 8 1 88%	8 - 13% - 88% - pos no
ponctuation_double	Doubled punctuation <i>je n'ai pas envie de tout refaire !!</i> <i>je n'ai pas envie de tout refaire !</i>	199 1.99% 50 3 94%	50 - - - 100% - pos yes
qu_elle_vs_quelle	Confusion of the homophones "qu'elle", "quel" and "quelle" <i>Qu'elle est la différence et peut on charger l'une ou l'autre?</i> <i>Quelle est la différence et peut on charger l'une ou l'autre?</i>	6 0.06% 6 - 100%	6 - - - 100% - pos yes
quelque_vs_quel_que	Confusion of the homophones "quelque", "quel que" and "quelle que" <i>La licence reste "attaché" à un utilisateur, quelque soit le PC.</i> <i>La licencereste "attaché" à un utilisateur, quel que soit le PC.</i>	0 0	- - - - -
soit_vs_sois_vs_so	Confusion of the homophones "soit", "sois" and "soi" <i>il ne peut plus faire de sauvegarde soit disant que mon Backup serait configuré à 0</i> <i>il ne peut plus faire de sauvegarde soi disant que mon Backup serait configuré à 0</i>	2 0.02% 2 - 100%	2 - - - 100% - pos no

sur_vs_sur	Confusion of the homophones "sur" and "sûr" Et je suis <i>sur</i> d'avoir vu au début : Et je suis <i>sûr</i> d'avoir vu au début :			9	-	-	11%	89%	-	pos	yes		
tes_vs_t_es	Confusion of the homophones "tes" and "t'es" <i>tous</i> ceci me fait un peut peur <i>tout</i> ceci me fait un peut peur			2	-	-	50%	-	50%				
tous_vs_tout	Confusion of the homophones "tous" and "tout" <i>tous</i> ceci me fait un peut peur <i>tout</i> ceci me fait un peut peur			13	38%	8%	-	54%	-	pos	no		
Total	4247	42%	611	52	91%	561	24%	11%	9%	53%	4%	pos	yes

Symantec Set 2 (for humans, multiple suggestions, manual application)

	Flags	Rule Precision	Rule Impact	
	Total flags on 10k sentences % of rule application on 10k sentences	Evaluated Flags Wrong Flags Precision	Evaluated Flags No impact on translation Negative impact About the same Positive impact Unclear (no majority) Rule impact	Statistically significant p-value <0.05
accord_phrase_nominale	Agreement error in noun phrase Bonjour, <i>quel version</i> de norton 360 avez-vous ? Bonjour, <i>quelle version</i> de norton 360 avez-vous ? 196 1.96%	50 12 76%	50 6% pos	yes
accord_sujet_verbe	Error in subject-verb agreement Tout d'abord <i>je vous fait</i> part de ma configuration: Tout d'abord <i>je vous fais</i> part de ma configuration: 237 2.37%	50 8 84%	50 2% pos	no
et_vs_est	Confusion of the homophones "et" and "est" Le problème <i>et</i> que je ne trouve pas l'endroit ou il faut entrée le code. Le problème <i>est</i> que je ne trouve pas l'endroit ou il faut entrée le code. 0		0	
evitez_est_ce_que	Use inversion instead of "est-ce que" for questions <i>Est ce que je dois</i> désinstaller et réinstaller Norton après coup ou pas? <i>Dois-je</i> désinstaller et réinstaller Norton après coup ou pas? 55 0.55%	50 3 94%	50 4% pos	no
evitez_le_langage_familier	Avoid colloquial language Voilà deux jours que <i>je galère</i> à essayer de créer ce SRD personnalisé sans succès. Voilà deux jours que <i>je me bats</i> à essayer de créer ce SRD personnalisé sans succès. 23 0.23%	23 - 100%	23 - pos	yes
evitez_le_participe_present	Avoid present participle Effectuez un test avec une autre suite de sécurité <i>ayant</i> les mêmes fonctionnalités que NIS Effectuez un test avec une autre suite de sécurité <i>qui a</i> les mêmes fonctionnalités que NIS ? 94 0.94%	50 2 96%	50 8% pos	yes
evitez_le_pluriel_entre_parenthe	Avoid using "(s)" for optional plurals <i>Élément(s)</i> de données du Registre <i>infecté(s)</i> : <i>Éléments</i> de données du Registre <i>infectés</i> : 9 0.09%	9 - 100%	9 11% pos	yes
evitez_les_anglicismes	Avoid anglicisms <i>je vais rebooter</i> et voir si c'est toujours le cas. <i>je vais redémarrer</i> et voir si c'est toujours le cas. 89 0.89%	50 10 80%	50 6% pos	no
evitez_les_phrases_clivees	Avoid cleft sentences <i>C'est Norton qui l'a supprimé ?</i> <i>Norton l'a supprimé ?</i> 35 0.35%	35 3 91%	32 6% pos	yes
evitez_les_questions_directes	Use inversion instead of direct questions <i>Tu as lu</i> le tuto sur le forum ? <i>As-tu lu</i> le tuto sur le forum ? 66 0.66%	50 - 100%	50 - pos	no
evitez_une_conjonction_en_deb	Remove conjunctions from BOS <i>Et</i> à chaque allumage d'ordinateur, j'ai droit à un grand écran de menace de suppression. À chaque allumage d'ordinateur, j'ai droit à un grand écran de menace de suppression. 396 3.96%	50 - 100%	50 - pos	yes

evitezAbrevForum	Avoid abbreviations <i>Je suis allé sur le site, j'ai téléchargé l'appli.</i> <i>Je suis allé sur le site, j'ai téléchargé l'application.</i>	417 4.17%	50 - 100%	50		pos	yes
fin_de_phrase_sans_ponctuation	End of sentence without punctuation <i>2- Désinstalle NIS avec Norton Removal Tool Tu télécharges et tu l'exécutes.</i> <i>2- Désinstalle NIS avec Norton Removal Tool. Tu télécharges et tu l'exécutes.</i>	96 0.96%	50 6 88%	50		pos	yes
forme_verbale_incorrecte	Wrong verb form <i>Et aussi, quand que tu avais installer norton, as tu eu des problemes d'installations?</i> <i>Et aussi, quand que tu avais installé norton, as tu eu des problemes d'installations?</i>	158 1.58%	50 4 92%	50		pos	yes
homophones_verbe_nom	Confusion of verb/noun homophones <i>Quand je clie sur ton lien j'arrive à une page qui m'indique norton internet security 2011</i> <i>Quand je clique sur ton lien j'arrive à une page qui m'indique norton internet security 2011</i>	27 0.27%	27 7 74%	20		pos	no
merci_de_tenir_ac	Incomplete expression "merci de tenir" <i>Merci de tenir au courant</i> <i>Merci de nous tenir au courant</i>	12 0.12%	12 - 100%	12		pos	yes
ne_pas_elider	Incorrect elision <i>Le "DiskPart" d' XP ne permet pas</i> <i>Le "DiskPart" de XP ne permet pas</i>	5 0.05%	5 - 100%	5		pos	no
negation_incomplete	Incomplete negation <i>je vais relire ce que tu m'as dit sur les antivirus pour être sure de pas avoir de problèmes!</i> <i>je vais relire ce que tu m'as dit sur les antivirus pour être sure de ne pas avoir de problèmes!</i>	180 1.80%	50 4 92%	50		pos	yes
passe_compose_avec_etre	Wrong auxiliary verb <i>Merci d'avoir revenu sur le forum et de nous avoir tenu au courant de la situation...</i> <i>Merci d'être revenu sur le forum et de nous avoir tenu au courant de la situation...</i>	0					
sequence_incorrecte_de_mots	Incorrect word sequence <i>Il suffit de démarre ton PC avec le CD (attention , logiciel en anglais , mais simple d'emploi).</i> <i>Il suffit de démarrer ton PC avec le CD (attention , logiciel en anglais , mais simple d'emploi).</i>	76 0.76%	50 17 66%	50		pos	yes
utilisez_le_subjonctif	Use subjunctive <i>Bien que je ne comprends pas ce que tu as fait quand tu écris :</i> <i>Bien que je ne comprenne pas ce que tu as fait quand tu écris :</i>	29 0.29%	29 8 72%	21		neg	no
wordDotWord	Missing space <i>je pense que je n'ai pas dû activer la clé du produit lors de mon reformatage.comment le faire?</i> <i>je pense que je n'ai pas dû activer la clé du produit lors de mon reformatage. Comment le faire?</i>	3 0.03%	3 1 67%	2		pos	no
autoSuggest_Bon	Remove "bon" from BOS <i>Bon, je viens de finir avec le support technique.</i> <i>Je viens de finir avec le support technique.</i>	15 0.15%	15 - 100%	15		pos	yes
autoSuggest_evitez_expressions	Replace expression "quand même" <i>Tu as quand même fait ta mise à jour en version 2012 comme je l'ai expliqué plus haut ?</i> <i>Tu as pourtant fait ta mise à jour en version 2012 comme je l'ai expliqué plus haut ?</i>	143 1.43%	50 - 100%	50		pos	yes

Symantec Set 3 (for the machine, unique suggestions, automatic application)

	Flags	Rule Precision	Rule Impact					
	Total flags on 10k sentences % of rule application on 10k sentences	Evaluated Flags Wrong Flags Precision	Evaluated Flags No impact on translation Negative impact About the same Positive impact Unclear (no majority) Rule impact Statistically significant					
ajoutez_dois_je	Add "dois-je" in questions with infinitive <i>Comment faire pour en supprimer une?</i> <i>Comment dois-je faire pour en supprimer une?</i> 36 0.36%	n.a.	36 6% 6% - 78% 11% pos	yes				
autoSuggest_abreviationIncorrec	Use of an incorrect abbreviation <i>J'ai un 2e PC installé de la même façon et sur celui-ci, je la trouve sans problème.</i> <i>J'ai un 2ème PC installé de la même façon et sur celui-ci, je la trouve sans problème.</i> 16 0.16%	n.a.	16 13% - 19% 69% 0% pos	yes				
autoSuggest_avoir_beau	Replace expression "avoir beau" <i>J'ai beau cliquer sur les deux pour lancer NIS, rien ne se passe encore une fois.</i> <i>Même si je clique sur les deux pour lancer NIS, rien ne se passe encore une fois.</i> 3 0.03%	n.a.	3 - - - 100% 0% pos	no				
autoSuggest_evitezMerciDe	Replace expression "merci de [...] tenir" <i>Merci de nous tenir au courant</i> <i>Veuillez nous tenir au courant</i> 21 0.21%	n.a.	21 5% 19% - 76% 0% pos	yes				
autoSuggest_formules_politesse	Replace greetings by standard expressions <i>Hello ElChikito, et bienvenue sur le Forum!!</i> <i>Bonjour ElChikito, et bienvenue sur le Forum!!</i> 19 0.19%	n.a.	19 16% 16% 26% 32% 11% pos	no				
autoSuggest_il_faut_que	Do not use "il faut que" <i>Il faut que je vérifie à nouveau.</i> <i>Je dois vérifier à nouveau.</i> 10 0.10%	n.a.	10 - 20% - 70% 10% pos	no				
autoSuggest_langage_familier	Remove or replace colloquial language <i>Oups désolé, j'ai oublié de préciser que j'ai NIS 2011</i> <i>Désolé, j'ai oublié de préciser que j'ai NIS 2011</i> 18 0.18%	n.a.	18 11% 44% 11% 28% 6% neg	no				
autoSuggest_ne_manquez_pas	Replace expression "ne manquez pas de" <i>Bien entendu, ne manquez pas de me tenir informé.</i> <i>Bien entendu, merci de me tenir informé.</i> 1 0.01%	n.a.	1 100% - - - 0%					
autoSuggest_tout	Move "tout" after verb <i>Si tu as installé NAV et que tu as tout laissé par défaut, je ne comprends pas.</i> <i>Si tu as installé NAV et que tu as laissé tout par défaut, je ne comprends pas.</i> 23 0.23%	n.a.	23 30% 17% - 52% - pos	no				
autoSuggest_utilisez_seulement	Use "seulement" instead of "ne que" <i>Je n'ai que la version d'évaluation.</i> <i>J'ai seulement la version d'évaluation.</i> 30 0.30%	n.a.	30 17% 10% - 70% 3% pos	yes				
autoSuggest_utilisezCa	Replace clitics by "ça" <i>Je te le donne en pièce jointe.</i> <i>Je te donne ça en pièce jointe.</i> 557 5.57%	n.a.	50 22% 6% 8% 58% 6% pos	yes				

autoSuggest_utilisezCeuxCi	Replace plural clitics by "ceux-ci" Je suis un fidèle des produits de la Symantec et les conseille autour de moi. Je suis un fidèle des produits de la Symantec et conseille ceux-ci autour de moi.	69 0.69%	n.a.	50		2%	pos	yes
évitez_jamais_après_verbe	Move "jamais" in front of verb to match English word order (also removes "ne") Si Norton a été désinstallé et que tu n'avais jamais fait de sauvegarde des identifiants, c'est perdu. Si Norton a été désinstallé et que tu jamais avais fait de sauvegarde des identifiants, c'est perdu.	8 0.08%	n.a.	8		-	neg	no
évitez_me_m_a	Remove or replace clitics Vous pouvez cliquer sur le lien qui vous renverra à votre sujet. Vous pouvez cliquer sur le lien qui renverra à votre sujet.	580 5.80%	n.a.	50		8%	pos	no
évitez_rien_avant_infinitif	Move "rien" after infinitive to match English word order Comment faire une image à froid avec le Symantec Recovery Disk' sans rien modifier ? Comment faire une image à froid avec le Symantec Recovery Disk' sans modifier rien ?	40 0.40%	n.a.	40		8%	pos	no
evitez_tu	Replace "tu" by "vous" Merci de ton aide! Merci de votre aide!	1968 20%	n.a.	50		2%	pos	yes
evitez_verbe_plus_rien	Remove "plus" before "rien" or "aucun" Non, il n'y a plus aucun Norton installé. Non, il n'y a aucun Norton installé.	23 0.23%	n.a.	23		0%	pos	no
évitez_verbe_plus_infinitif	Avoid Verb followed by infinitive Je pense avoir trouvé la solution à votre problème sur le Forum US de Norton. Je pense que j'ai trouvé la solution à votre problème sur le Forum US de Norton.	not evaluated	n.a.	not evaluated				
		3422 34%	n.a.	448		5%	pos	yes

TSF Set 1 (for humans, unique and multiple suggestions, manual application)

	Flags	Rule Precision	Rule Impact	
	Total flags on 3k sentences % of rule application on 3k sentences	Evaluated Flags Wrong Flags Precision	Evaluated Flags No impact on translation Negative impact About the same Positive impact Unclear (no majority)	Rule impact Statistically significant
				p-value <0.05
a_vs_a	Confusion of the homophones "a", "as" and "à" A l'exception de Buram, les faits de guerre (violence) ne sont pas les plus importants. À l'exception de Buram, les faits de guerre (violence) ne sont pas les plus importants. 33 1.32%	33 - 100%	32 13% 6% 6% 63% 13%	pos yes
accord_phrase_nominale	Agreement error in noun phrase Mis à part les 3 médecins expatriés le seul médecins rencontré est le Directeur Général du Secrétariat de la Santé. Mis à part les 3 médecins expatriés le seul médecin rencontré est le Directeur Général du Secrétariat de la Santé. 93 3.72%	93 31 67%	50 36% 32% 4% 22% 6%	neg no
accord_sujet_verbe	Error in subject-verb agreement Un deuxième médecin expatrié n'aurai pas été présent lors de l'évènement. Un deuxième médecin expatrié n'aurait pas été présent lors de l'évènement. 18 0.72%	18 11 39%	7 - 14% - 86% -	pos no
ajoutez_un_blanc	Insert a space 3 Rapports mensuels du Creni Magaria oct. 2010 3 Rapports mensuels du Creni Magaria oct. 2010 19 0.76%	19 - 100%	19 42% - 21% 32% 5%	pos yes
ajoutez_un_trait_d_union	Missing hyphenation c'est à dire ? c'est - à - dire ? 12 0.48%	12 - 100%	12 17% 25% - 50% 8%	pos no
ajoutez_une_virgule	Missing comma De plus l'existence du fond permet d'éviter aux familles de choisir entre vendre tous leurs moyens de subsistance. De plus, l'existence du fond permet d'éviter aux familles de choisir entre vendre tous leurs moyens de subsistance. 45 1.80%	45 - 100%	45 13% 11% - 69% 7%	pos yes
des_vs_des	Confusion of the homophones "des", "dès" and "dés" Dés qu'un système fiable d'assurance santé fonctionnera, ce projet pilote à base communautaire devra s'arrêter. Dès qu'un système fiable d'assurance santé fonctionnera, ce projet pilote à base communautaire devra s'arrêter. 11 0.44%	11 9 18%	2 - 50% - 50% -	neg no
du_vs_du	Confusion of the homophones "du" and "dû" nombre de femmes accouchaient en route, une césarienne a du être pratiquée en catastrophe. nombre de femmes accouchaient en route, une césarienne a dû être pratiquée en catastrophe. 1 0.04%	1 - 100%	1 - - - 100% -	pos no
elidez_ce_mot	Missing or incorrect elision Le plus important était que à partir de maintenant elles seront partie intégrante du Conseil Général International. Le plus important était qu'à partir de maintenant elles seront partie intégrante du Conseil Général International. 1 0.04%	1 - 100%	1 100% - - -	0 0
erreur_de_majuscule	Wrong capitalisation Les dépenses du gouvernement pour la santé ne représentaient en 2009 que 0,2% du Produit intérieur brut. Les dépenses du gouvernement pour la santé ne représentaient en 2009 que 0,2% du produit intérieur brut. 143 5.72%	143 24 83%	50 20% 18% 12% 46% 4%	pos yes
espaces_autour_ponctuation	Incorrect spaces before or after punctuation Il s'agit d'intervenir sur plusieurs causes de la surmortalité infantile: Il s'agit d'intervenir sur plusieurs causes de la surmortalité infantile : 121 4.84%	121 - 100%	50 100% - - -	0 0
évitiez_adverbe	Avoid adverb A-t-on vraiment besoin de parler des 30 ans d'MSF Suisse ? A-t-on besoin de parler des 30 ans d'MSF Suisse ? 5 0.20%	5 1 80%	4 25% 25% - 50% -	pos no

evitez_les_anglicismes	Avoid anglicisms <i>Exercice de management du Shared Fund (2h)</i> <i>Exercice de gestion du Shared Fund (2h)</i> 8 0.32% 8 1 88% 7 57% 14% - 29% - pos no
evitez_ce_qui	Avoid "ce qui" <i>En effet, il n'existe aucun plan préfectoral de suivi, ce qui se traduit par:</i> <i>En effet, il n'existe aucun plan préfectoral de suivi. Ceci se traduit par:</i> 6 0.24% 6 - 100% 6 - 17% - 67% 17% pos no
evitez_une_conjonction_en_debut	Remove conjunctions from BOS <i>Mais la qualité n'est pas non plus assurée par manque de visibilité et de supervision des activités.</i> <i>La qualité n'est pas non plus assurée par manque de visibilité et de supervision des activités.</i> 15 0.60% 15 2 87% 13 - - - 100% - pos yes
evitez_incise_apres_qui	Avoid interpolated clauses introduced by "qui" <i>Ce n'est pas forcément le cas pour les villageois qui, ayant contribué, s'attendent à recevoir un support.</i> <i>Ce n'est pas forcément le cas pour les villageois. Puisqu'ils ont contribué, ils s'attendent à recevoir un support.</i> 1 0.04% 1 - 100% 1 - - - 100% - pos no
evitez_le_participe_present	Avoid present participle <i>90 % des cas étant du paludisme simple</i> <i>comme 90 % des cas sont du paludisme simple</i> 120 4.80% 120 - 100% 50 10% 8% 2% 74% 6% pos yes
evitez_ponctuation	Unnecessary punctuation <i>Depuis 2004, plaidoyer commun sur la prise en charge médicale pour les étrangers / Sans-papiers, (projet HUMA)</i> <i>Depuis 2004, plaidoyer commun sur la prise en charge médicale pour les étrangers / Sans-papiers (projet HUMA)</i> 59 2.36% 59 - 100% 50 2% 32% 4% 34% 28% pos no
expression_incorrecte	Spelling error in common expression <i>Quand on m'a proposé de poser ma candidature au CA, j'ai dit " oui, bien sur !</i> <i>Quand on m'a proposé de poser ma candidature au CA, j'ai dit " oui, bien sûr !</i> 2 0.08% 2 1 50% 1 100% - - - - 0 0
forme_verbale_incorrecte	Wrong verb form <i>Et aussi, quand que tu avais installer norton, as tu eu des problemes d'installations?</i> <i>Et aussi, quand que tu avais installé norton, as tu eu des problemes d'installations?</i> 8 0.32% 8 2 75% 6 17% 17% - 33% 33% pos no
homophones_divers	Confusion of homophones - others <i>Le diagnostique, la référence et le traitement (complications incluses) sont disponibles gratuitement.</i> <i>Le diagnostic, la référence et le traitement (complications incluses) sont disponibles gratuitement.</i> 4 0.16% 4 1 75% 3 33% 33% - - 33% neg no
homophones_verbe_nom	Confusion of verb/noun homophones <i>la charge de travaille d'août à novembre fut surélevée, doubles journées de travail n'étant pas extraordinaires.</i> <i>la charge de travail d'août à novembre fut surélevée, doubles journées de travail n'étant pas extraordinaires.</i> 3 0.12% 3 - 100% 3 - - - 100% - pos no
evitez_lettre_entre_parentheses	Avoid using "(s)" for optional plurals <i>Offensive envisagée (s) sur le corridor entre mars et la saison des pluies.</i> <i>Offensive envisagée sur le corridor entre mars et la saison des pluies.</i> 7 0.28% 7 3 57% 3 - - - 100% - pos no
evitez_phrase_en_majuscules	Avoid writing entire sentence in capital letters <i>SCHÉMA RECAPITULATIF DES ACTIVITES MISES EN PLACE DEPUIS 2009</i> <i>Schéma récapitulatif des activités mises en place depuis 2009</i> 122 4.88% 122 9 93% 50 18% 2% - 80% - pos yes
evitez_les_phrases_clivees	Avoid cleft sentences <i>Préciser qui est le team leader de l'équipe et quelles sont les compétences des membres de l'équipe.</i> <i>Préciser qui est le team leader de l'équipe et les compétences des membres de l'équipe.</i> 4 0.16% 4 - 100% 4 - - - 100% - pos no

evitez_les_questions_directes	Use inversion instead of direct questions <i>Vous avez un trou de mémoire ?</i> <i>Avez-vous un trou de mémoire ?</i>			1	-	-	-	100%	-	pos	no					
sequence_incorrecte_de_mots	Incorrect word sequence <i>La population était estimée en 2008 à 329 455 personnes avec une la distribution géographique suivante :</i> <i>La population était estimée en 2008 à 329 455 personnes avec la distribution géographique suivante :</i>			9	0.36%	9	2	78%	7	29%	-	-	71%	-	pos	no
utilisez_de	Use "de" instead of "des" <i>pourquoi ne pas, dès lors, proposer que l'un des ces 2 postes soit alloué à quelqu'un du terrain ?</i> <i>pourquoi ne pas, dès lors, proposer que l'un de ces 2 postes soit alloué à quelqu'un du terrain ?</i>			3	0.12%	3	1	67%	2	-	50%	-	50%	-	neg	no
verbe_sujet_inverse	Move the subject in front of the verb <i>En résumé, les principaux problèmes/défis que compte le pays dans le domaine de la santé sont les suivants :</i> <i>En résumé, les principaux problèmes/défis que le pays compte dans le domaine de la santé sont les suivants :</i>			16	0.64%	16	10	38%	6	-	50%	-	50%	-	neg	no
Total	890	36%	890	95	89%	486	26%	13%	3%	49%	10%	pos	yes			

TSF Set 2 (for the machine, unique suggestions, automatic application)

	Flags		Rule Precision			Rule Impact							Statistically significant
	Total flags on 3k sentences	% of rule application on 3k sentences	Evaluated Flags	Wrong Flags	Precision	Evaluated Flags	No impact on translation	Negative impact	About the same	Positive impact	Unclear (no majority)	Rule impact	
autoSuggest_abreviationIncorrecte	<p>Use of an incorrect abbreviation</p> <p>Préciser les périodes auxquelles la mission peut prendre place (ex/ 1er jour de démarrage possible).</p> <p>Préciser les périodes auxquelles la mission peut prendre place (ex/ premier jour de démarrage possible).</p>												
	7	0.28%	7	-	100%	7	29%	-	-	71%	-	pos	no
autoSuggest_ajouter_une_virgule_a_près_PP	<p>Add a comma after a prepositional phrase</p> <p>Depuis le mois de juin seul l'Hôpital de Gedel, appuyé par le Diocèse of El Obeid</p> <p>Depuis le mois de juin, seul l'Hôpital de Gedel, appuyé par le Diocèse of El Obeid</p>												
	23	0.92%	23	-	100%	23	17%	4%	13%	61%	4%	pos	yes
autoSuggest_évitéz_barre_oblique	<p>Do not use slash</p> <p>Un brainstorming a été fait le dernier jour pour définir une / des priorités et planifier la semaine.</p> <p>Un brainstorming a été fait le dernier jour pour définir des priorités et planifier la semaine.</p>												
	1	0.04%	1	-	100%	1	-	-	-	100%	-	pos	no
autoSuggest_évitéz_en	<p>Do not use the pronoun "en"</p> <p>Une formation continue a été mise en place avec des séances de rattrapage pour celles qui en avaient besoin.</p> <p>Une formation continue a été mise en place avec des séances de rattrapage pour celles qui avaient besoin de ça.</p>												
	3	0.12%	3	-	100%	3	-	100%	-	-	-	neg	no
autoSuggest_évitézExpressions_TSF	<p>Replace expressions like "de ce fait"</p> <p>De ce fait, la majorité de la population n'a pas les moyens de se faire soigner.</p> <p>Pour cette raison, la majorité de la population n'a pas les moyens de se faire soigner.</p>												
	31	1.24%	31	-	100%	31	23%	16%	-	58%	3%	pos	yes
autoSuggest_nounAdj	<p>Place the adjective before the noun it modifies</p> <p>Le groupe géopolitique présente une stratégie par continent aux directeurs.</p> <p>Le géopolitique groupe présente une stratégie par continent aux directeurs.</p>												
	1168	46.72%	1168	5	100%	50	40%	20%	2%	28%	10%	pos	no
autoSuggest_nounAdjAdj	<p>If there are two adjective, place the last one before the noun</p> <p>Les infections respiratoires sévères sont la première cause de consultation.</p> <p>Les sévères infections respiratoires sont la première cause de consultation.</p>												
	71	2.84%	71	5	93%	50	14%	24%	-	58%	4%	pos	yes
autoSuggest_nounAdjetAdj	<p>Place adjectives joined by a conjunction before the noun they modify</p> <p>Ce serait utile pour différencier une hématurie chez les malades atteints d'une affection néphrologique ou urologique.</p> <p>Ce serait utile pour différencier une hématurie chez les malades atteints d'une néphrologique ou urologique affection.</p>												
	31	1.24%	31	1	97%	31	13%	23%	-	58%	6%	pos	yes
autoSuggest_nounProperNoun	<p>Place the proper noun before the head</p> <p>En ce qui concerne la mise en œuvre du projet Guékédou, une collaboration avec les différents acteurs est nécessaire.</p> <p>En ce qui concerne la mise en œuvre du Guékédou projet, une collaboration avec les différents acteurs est nécessaire.</p>												
	38	1.52%	38	1	97%	38	-	34%	3%	58%	5%	pos	no
autoSuggest_nounSigle	<p>Place the acronym before the noun it refers to</p> <p>Ce résident pourrait aussi bénéficier de la formation délivrée par le médecin MSF.</p> <p>Ce résident pourrait aussi bénéficier de la formation délivrée par le MSF médecin.</p>												
	62	2.48%	62	-	100%	50	20%	8%	-	70%	2%	pos	yes
autoSuggest_passéSimple	<p>Avoid the simple past tense</p> <p>Les médecins nigériens se sentirent trahis.</p> <p>Les médecins nigériens se sentaient trahis.</p>												
	25	1.00%	25	-	100%	25	16%	24%	8%	52%	-	pos	no

p-value <0.05

autoSuggest_utilisez_seulement	Use "seulement" instead of "ne que" Cela représente une forte croissance (10%) puisque nous <i>n' avons que</i> 3312 employés locaux en 2010. Cela représente une forte croissance (10%) puisque nous avons <i>seulement</i> 3312 employés locaux en 2010.		3	0.12%	3	-	100%	3	-	67%	-	33%	-	neg	no
autoSuggest_utilisezÇa_TSF	Replace clitics by "ça" Les données statistiques <i>le</i> font ressortir comme première cause de consultations (33,8%). Les données statistiques font ressortir <i>ça</i> comme première cause de consultations (33,8%).		24	0.96%	24	5	79%	24	8%	25%	4%	54%	8%	pos	no
autoSuggest_utilisezCeuxCi_TSF	Replace plural clitics by "ceux-ci" Aucune notification ne <i>les</i> différencie des décès causés par le manque de sang. Aucune notification ne différencie <i>ceux-ci</i> des décès causés par le manque de sang.		25	1.00%	25	1	96%	25	24%	20%	-	52%	4%	pos	no
évitez_dont	Avoid the relative pronoun "dont" 250 personnes <i>dont</i> 50 enfants. 250 personnes <i>parmi lesquels</i> 50 enfants.		20	0.80%	20	-	100%	20	30%	5%	5%	50%	10%	pos	yes
évitez_tu_TSF	Replace "tu" by "vous" Et <i>toi</i> , quelles réponses <i>apporterais-tu</i> aux questions qui sont posées ? Et <i>vous</i> , quelles réponses <i>apporteriez-vous</i> aux questions qui sont posées ?		9	0.36%	9	3	67%	9	22%	11%	11%	33%	22%	pos	no
évitez_préfixeRe	Replace the prefix "re" (or "ré") by "à nouveau" Nous devons aujourd'hui <i>re-</i> dynamiser notre capacité d'innovation. Nous devons aujourd'hui dynamiser <i>à nouveau</i> notre capacité d'innovation.		1	0.04%	1	-	100%	1	-	-	-	100%	-	pos	no
Total	1543	62%	1543	21	99%	392	19%	20%	3%	54%	5%	pos	yes		

Appendix 2: List of English Pre-editing Rules

This appendix lists and describes all English pre-editing rules in alphabetical order.

a_an_distinction

Short Description

General Guidelines

Pay attention to the sound that the word begins with, not necessarily the letter. Use a before a word that begins with a consonant sound. Use an before a word that begins with a vowel sound.

- an interface module
- an x-ray machine
- a user interface
- a new user

Variation: Acronyms

Acronyms can cause difficulty when deciding between a and an. Think about how you would say the acronym aloud. Do you pronounce each letter individually, or do you say it like a word?

- an RSS news feed
- an SEC ruling
- a RAM module
- a NATO conference

Some acronyms are commonly pronounced both ways. Consult the best practice guidelines of your company to decide which variant to use.

- an FAQ forum
- a FAQ forum (pronounced *fak*)
- an SQL server
- a SQL server (pronounced *sequel*)

Variation: Regional and stylistic variation

Some words are pronounced differently depending on region or style. These words, especially those beginning with h, can vary in the choice of a or an.

- an herb
- a herb
- an historic occasion
- a historic occasion

Examples

- Create a folder.
- Create an folder.
- Services are billed at an hourly rate.
- Services are billed at a hourly rate.
- Installing an additional plug-in is quick and easy.
- Installing a additional plug-in is quick and easy.
- Follow these steps to modify a user interface.

- Follow these steps to modify an user interface.
- [Click here to view the file as an HTML document.](#)
- [Click here to view the file as a HTML document.](#)

adjective_adverb_confusion

Short Description

Use adjectives to describe nouns.

Use adverbs to describe verbs, adjectives, or other adverbs.

Examples

- The system has correctly recognized the issue.
- The system has correct recognized the issue.
- This software has effectively corrected the errors.
- This software has effective corrected the errors.
- That robot was intensively trained.
- That robot was intensive trained.
- The guests are greeted by the host in a friendly way.
- The guests are friendly greeted by the host.

affect_effect_impact_confusion

Short Description

The words affect, effect, and impact can cause confusion.

Use **affect** as a verb, meaning to influence.

- Delays affect the company's bottom line.
- See how the changes affect you.

Use **effect** as a noun, meaning a result or change.

- Delays have an effect on the company's bottom line.
- See what effect the changes have on you.

Note:effect can also be used as a verb, meaning to bring about.

Use **impact** as a noun to express a strong effect. Do not use **impact** as a verb.

- Delays have an impact on the company's bottom line.
- See what impact the changes have on you.

Examples

- The software does not affect system performance.
- The software does not effect system performance.
- Elevated blood pressure is one effect of the medication.
- Elevated blood pressure is one affect of the medication.
- The new interstate has an impact on the local community.
- The new interstate impacts the local community.

amount_number_confusion

Short Description

Use **number** with nouns that can be counted, such as *document* or *customer*.

- number of documents
- number of customers

Use **amount** with nouns that cannot be counted, such as *interest* or *information*.

- amount of interest
- amount of information

Examples

- The region has a substantial amount of annual rainfall.
- The region has a substantial number of annual rainfall.
- In the airline industry an overwhelming amount of information is collected each year.
- In the airline industry an overwhelming number of information is collected each year.
- The document contained a large number of mistakes.
- The document contained a large amount of mistakes.

avoid_colloquialism_and_metaphorical_language

Short Description

Colloquialisms are informal expressions that are difficult to translate automatically, as well as metaphorical language.

Examples

- They had problems from the beginning.
- They had problems right off the bat.
- In summary, the meeting went well.
- All in all the meeting went well.

avoid_duplicates

Short Description

Avoid duplicate words that make the meaning nonsensical. Such mistakes often result when a writer rewords a phrase while typing but does not recheck the entire sentence for clarity.

Examples

- There are two rules that could be both true.
- There are two rules that could be both be true.
- The dog is barking.
- The a dog is barking.

- Tighten the screws to the left.
- Tighten the screws to the on the left.
- We have many processes that depend on support cases being tracked in this system.
- We have some many processes that depend on support cases being tracked in this system.
- The command buttons have to be touched in the release or execute mode.
- The command buttons have to be to touched in the release or execute mode.

capitalize_at_beginning_of_sentence

Short Description

Capitalize the first letter of a word that begins a sentence.

Examples

- This is the definition.
- this is the definition.

check_possible_wrong_suffix

Short Description

Words ending in *-ency* or *-ence*, *-ancy* or *-ance*, *-ic* or *-ical* are easily confused. In many cases, one form is much more common than the other.

Rare form	Common form
<i>consistence</i>	<i>consistency</i>
<i>acceptancy</i>	<i>acceptance</i>
<i>technic</i>	<i>technical</i>

In other cases, the forms have different meanings and are only appropriate for use in certain contexts.

Rare use	Common use
<i>economical growth</i>	<i>economic growth</i>
<i>economical policy</i>	<i>economic policy</i>
<i>economic car</i>	<i>economical car</i>
<i>economic heating</i>	<i>economical heating</i>

Note: Terms entered in the Acrolinx terminology are treated as exceptions to this rule.

Examples

- The frequency operation requires more input.

- The frequency operation requires more input.
- Carry out a redundancy test.
- Carry out a redundancy test.
- The economic recovery was slower than expected.
- The economical recovery was slower than expected.
- Low-cost carriers offer tickets at an economical price.
- Low-cost carriers offer tickets at an economic price.

do_not_use_this_word

Short Description

Do not use the flagged word under any circumstances.

do_not_use_whether_or_not

Short Description

Do not use *or not* with *whether* to indicate a choice between alternatives. It is redundant, because by using *whether* you have already communicated the notion of choice.

Examples

- To show whether the financial instrument is an odd coupon, use the code ODDC.
- To show whether or not the financial instrument is an odd coupon, use the code ODDC.
- The only parameter that you can modify is the Distribution field, which specifies whether you want to log the event or ignore it.
- The only parameter that you can modify is the Distribution field, which specifies whether or not you want to log the event or ignore it.

duplicate_punctuation_mark

Short Description

Avoid duplicate punctuation marks.

Examples

- The following example can be used to demonstrate how Acrolinx deals with lists:
- The following example can be used to demonstrate how Acrolinx deals with lists::
- If we spent more time in Germany, we would understand the language better.
- If we spent more time in Germany,, we would understand the language better.
- Label the package as fragile when packing for transport.
- Label the package as fragile when packing for transport..

fewer_less_confusion

Short Description

Use **fewer** with nouns that can be counted, such as *document* or *customer*.

- fewer documents
- fewer customers

Use **less** with nouns that cannot be counted, such as *interest* or *information*.

- less interest
- less information

Examples

- After archiving, you can search the Event Journal more quickly, because it contains fewer events.
- After archiving, you can search the Event Journal more quickly, because it contains less events.
- Increasing this value means that the server spends less time with the setup.
- Increasing this value means that the server spends fewer time with the setup.
- Perhaps within nine months, or in even less time, we will see things accelerating again.
- Perhaps within nine months, or in even fewer time, we will see things accelerating again.
- The situation of young people with fewer opportunities is an issue of concern in many European countries.
- The situation of young people with less opportunities is an issue of concern in many European countries.

incorrect_extra_comma

Short Description

Commas in English are generally not used before *that*, *in order to*, *so that*, *if*, or time and place words such as *when* and *where*.

Examples

- Make sure that the setting is switched on.
- Make sure, that the setting is switched on.
- You need to press enter in order to complete setup.
- You need to press enter, in order to complete setup.
- Switch on the screen so that you can see warning messages.
- Switch on the screen, so that you can see warning messages.
- Switch on the screen if you want to see warning messages.
- Switch on the screen, if you want to see warning messages.
- Start the program when you have installed all the necessary components.
- Start the program, when you have installed all the necessary components.

incorrect_preposition

Short Description

Make sure to use the correct preposition.

- demand for
- effect on
- equivalent to
- identical to
- independent of
- participate in
- mean a lot to
- order by
- depending on
- different from
- similar to
- the same as
- supported by
- in my opinion

Make sure no preposition after the following verbs.

- discuss
- enter
- reach
- resemble

Examples

- You need to participate in this session.
- You need to participate to this session.
- She entered the room.
- She entered into the room.
- There is a growing demand for energy improvements.
- There is a growing demand after energy improvements.
- The ministers discuss the crisis.
- The ministers discuss about the crisis.
- The eurozone crisis has a effect on the economy.
- The eurozone crisis has a effect over the economy.
- The study says that smoking one joint is equivalent to 20 cigarettes.
- The study says that smoking one joint is equivalent with 20 cigarettes.
- Mental states are identical to brain states.
- Mental states are identical with brain states.
- The power supply is independent of the resistance.
- The power supply is independent on the resistance.
- This project means a lot to me.
- This project means a lot for me.
- We reached the airport at 9 pm.
- We reached at the airport at 9 pm.
- The child resembles its mother.
- The child resembles to its mother.

- Greece is depending on the financial aids from other European countries.
- Greece is depending after the financial aids from other European countries.
- The tax system in Germany is very different from the one in Greece.
- The tax system in Germany is very different with the one in Greece.
- Singapore is similar to Hong Kong.
- Singapore is similar with Hong Kong.
- The CEO is supported by all the staff.
- The CEO is supported with all the staff.
- In my opinion, I do not think it is feasible.
- To my opinion, I do not think it is feasible.
- A light beam can bridge the distance from one side of the street to the other.
- A light beam can bridge the distance to one side of the street to the other.

irregular_verb_use

Short Description

Most verbs form the past tense by adding *-ed* or *-d*, for example:

Present	Past	Past Participle
start	started	have started
initiate	initiated	have initiated

Irregular verbs have different forms. Some common irregular verbs are the following:

Present	Past	Past Participle
be	was, were	have been
become	became	have become
begin	began	have begun
bring	brought	have brought
build	built	have built
buy	bought	have bought
choose	chose	have chosen
come	came	have come
cost	cost	have cost
do	did	have done
drive	drove	have driven
give	gave	have given
go	went	have gone
have	had	have had
know	knew	have known
let	let	have let

make	made	have made
pay	paid	have paid
say	said	have said
see	saw	have seen
sell	sold	have sold
send	sent	have sent
speak	spoke	have spoken
write	wrote	have written

Examples

- They ate a delicious dinner at the new restaurant.
- They eated a delicious dinner at the new restaurant.
- The doctor has read important books.
- The doctor has readed important books.
- The doctor has read important books.
- The doctor has readed important books.
- The system administrator went to the server room.
- The system administrator goed to the server room.
- In sum, we have sold nearly €5 billion in enterprise value since 2003.
- In sum, we have selled nearly €5 billion in enterprise value since 2003.
- The application fee is not to be paid, however, until you have received an appropriate demand for payment from the patent office.
- The application fee is not to be payed, however, until you have received an appropriate demand for payment from the patent office.
- All the information required for the individual options was included on the proof of entitlement to vote, sent to electors before every ballot.
- All the information required for the individual options was included on the proof of entitlement to vote, sended to electors before every ballot.

its_it_is_confusion

Short Description

The words *it's* and *its* can cause confusion.

With nouns, 's expresses belonging.

- The IT staff constantly review the system and the system's settings.
- The representative introduced the company and the company's CEO.

It is an exception to this rule and takes only *s* without an apostrophe to express belonging.

Use **it's** as a contraction of *it is*.

- It's important that we finish on time.
- We will know if it's working.

Use **its** to mean *belonging to it*.

- The IT staff constantly review the system and its settings.
- The representative introduced the company and its CEO.

Examples

- It's time for dinner.
- Its time for dinner.
- It's however possible to carefully insert a 3 mm glass slide.
- Its however possible to carefully insert a 3 mm glass slide.
- The system checks its log file.
- The system checks it's log file.
- In the period under review pharmexx also expanded its customer portfolio further.
- In the period under review pharmexx also expanded it's customer portfolio further.

loose_lose_confusion

Short Description

The words *lose* and *loose* can cause confusion.

Use **lose** as a verb. Its meaning is the opposite of *win* or *keep*.

- If the team does not have a winning strategy, they are likely to lose.
- The rewards program will help us not lose customers.

Use **loose** as an adjective. Its meaning is the opposite of *tight*.

- If the bolts are loose, tighten them with a wrench.
- Loose teeth are a possible indication of gum disease.

Examples

- The system is dynamic, and individual components can gain or lose their position.
- The system is dynamic, and individual components can gain or loose their position.
- If the roots are exposed to heavy metals, the stem cells lose their activity.
- If the roots are exposed to heavy metals, the stem cells loose their activity.
- Sophisticated software confirms that no parts of the apparatus are loose.
- Sophisticated software confirms that no parts of the apparatus are lose.

missing_space

Short Description

Insert a space after the punctuation.

Examples

- This parameter is assigned to a variable. The data structure components have the same function as the components.
- This parameter is assigned to a variable. The data structure components have the same function as the components.
- In order to operate, some applications require certain versions of these releases.
- In order to operate, some applications require certain versions of these releases.
- There are three stages: analysis, implementation, and evaluation.
- There are three stages: analysis, implementation, and evaluation.

missing_word

Short Description

There seems to be a missing word here.

The reason can be:

- missing a subject
- missing a verb
- missing a preposition
- missing a relative pronoun

Examples

- Data is stored by an object that is encapsulated.
- Data is stored by an object is encapsulated.
- Prepare the deliverables according to the requirements.
- Prepare the deliverables according the requirements.
- It begins by introducing procedural languages, where the system uses CICS commands to request services.
- It begins by introducing procedural languages, where uses CICS commands to request services.

much_many_confusion

Short Description

Use **many** with nouns that can be counted, such as *document* or *customer*.

- many documents
- many customers

Use **much** with nouns that cannot be counted, such as *interest* or *information*.

- much interest
- much information

Examples

- Many companies hold online conferences and trainings.
- Much companies hold online conferences and trainings.
- We may not always know how much stress is too much.
- We may not always know how many stress is too many.
- The systems analyst determines how much information there is to store.
- The systems analyst determines how many information there is to store.
- This is because national penal laws can be interpreted in many ways.
- This is because national penal laws can be interpreted in much ways.

must_hyphenate

Short Description

Some compound words always require a hyphen. Examples of these compound words are the following:

- cross-check
- double-click
- fail-safe
- fine-tune
- jump-start
- in-house
- line-beam
- non-*adjective*
- shrink-mounted
- user-friendly
- walkie-talkie
- well-being
- twenty-two
- ninety-one

Examples

- Some mechanisms allow you to fine-tune the signal strength.
- Some mechanisms allow you to fine tune the signal strength.
- Non-native speakers do not understand this text.
- Non native speakers do not understand this text.
- Select the corresponding calculation and double-click it.
- Select the corresponding calculation and double click it.
- The company handles its customer support in-house.
- The company handles its customer support in house.
- On long flights, it is important to look after your well-being.
- On long flights, it is important to look after your well being.
- Lastly, in 2003, an outstanding donation by the Georgia O'Keeffe foundation brought the museum twenty-two original *prints by Alfred Stieglitz.
- Lastly, in 2003, an outstanding donation by the Georgia O'Keeffe foundation brought the museum twenty two original prints *by Alfred Stieglitz.

noun_adjective_confusion

Short Description

Follow these general guidelines for determining whether two words should be written separately or hyphenated:

- Nouns are generally written separately.
- Adjectives are generally hyphenated.

Example:

- **back office** is a noun: *The back office remains hidden from customers. Important operations take place in the back office.*
- **back-office** is an adjective: *Back-office operations are extremely important. Important operations are often back-office.*

Examples

- The server has to identify the end user.
- The server has to identify the end-user.
- The developer presented a high-level overview of the software.
- The developer presented a high level overview of the software.
- The customers run the tests in real time.
- The customers run the tests in real-time.
- The board voted to implement the single-window concept.
- The board voted to implement the single window concept.
- Make sure that the webpage is open in a single window.
- Make sure that the webpage is open in a single-window.

noun_adjective_verb_confusion

Short Description

Follow these general guidelines for determining whether two words should be written separately, together, or hyphenated:

- Verbs are generally written separately.
- Nouns and adjectives are generally written together or hyphenated.

Example:

- **back up** is a verb: *The system administrator backs up important files.*
- **backup** or **back-up** is an adjective or noun: *Click here to begin a backup of your system. Click here to create a back-up file.*

Examples

- How to decide what data to back up.
- How to decide what data to backup.
- A periodic backup is important.

- A periodic back up is important.
- Press the green button to start up your device.
- Press the green button to startup your device.
- Insert the startup disk to continue.
- Insert the start up disk to continue.
- Click here to set up your user profile.
- Click here to setup your user profile.
- The machine restarts when warm-up is complete.
- The machine restarts when warm up is complete.

np_number_agreement

Short Description

Make sure that the noun agrees with the word before it. Use singular nouns with singular articles and numbers. Use plural nouns with plural articles and numbers.

Use the following with singular nouns:

- a
- an
- this
- one

Use the following with plural nouns:

- these
- those
- several
- all numbers except *one*

Examples

- The development team has an idea for a new project.
- The development team has an ideas for a new project.
- We must present those issues that are of primary interest to the reader.
- We must present those issue that are of primary interest to the reader.
- There are several factors in these broad guidelines.
- There are several factor in these broad guidelines.
- The kit comes with eight cables for your use.
- The kit comes with eight cable for your use.
- This table shows you the current versions of our products and their compatibility with different releases of QuarkXPress *and Mac OS.
- This tables show you the current versions of our products and their compatibility with different releases of QuarkXPress *and Mac OS.

reformulation

Short Description

This rule reformulates words or phrases to improve the output of SMT systems. It has been developed within the ACCEPT research project for translating Symantec forum texts.

Since the rule may have a negative or questionable impact on the source (input) text, it should only be used automatically to prepare text for machine translation.

Examples

- You must remember that the software is not perfect.
- You have to remember that the software is not perfect.

reformulation_tw

Short Description

This rule reformulates words or phrases to improve the output of SMT systems. It has been developed within the ACCEPT research project for translating Translators Without Borders documents.

Since the rule may have a negative or questionable impact on the source (input) text, it should only be used automatically to prepare text for machine translation.

Examples

- Therefore it's not always the obvious cause.
- So it's not always the obvious cause.

repeated_word

Short Description

Make sure that the repeated word is intentional.

Examples

- In this section, you make sure that the insurance decision tree overrides the general decision tree.
- In this section, you make sure that that the insurance decision tree overrides the general decision tree.
- Avoid exposing your device to environments with excessive dust or moisture.
- Avoid exposing your device exposing it to environments with excessive dust or moisture.

spelling_error

Short Description

This is a spelling error.

Examples

- vice versa
- vise versa
- error message
- error massage
- sneak peek
- sneak peak

subject_verb_agreement

Short Description

General Guidelines

The subject of the sentence must agree with the verb in number. In other words, if the noun or pronoun which *does* something is plural, then the verb must be plural as well. Likewise if the noun or pronoun is singular, the verb must be singular too.

- The interface **module is** ready.
- The interface **modules are** ready.

Special Cases

Pronouns: words like *anyone* or *no one*

Some pronouns always take a singular verb, even when they seem to have a plural meaning, for example *anyone, anybody, everyone, everybody, someone, somebody, no one, nobody, each, neither, and either*.

- **Everyone is** ready to begin.
- **Nobody knows** the correct procedure.
- **Each** of the computers **needs** repair.
- Do you have a preference? **Either is** acceptable.

There and here

In sentences starting with *there* or *here*, the subject of the sentence comes after the verb.

- Here **are** two possible **explanations** for the slow download.
- **Is** there a **way** to deactivate the firewall?

Collective nouns

Some words look like they have a plural form but are in fact singular, such as *news* or *measles*. On the other hand, some singular words ending in *-s* do take a plural verb, such as *assets, earnings, or thanks*. Consult a dictionary if you are unsure whether a word is singular or plural.

- The **news** from abroad **is** positive.
- The **earnings** from last quarter **are** encouraging.

Some collective nouns can be singular or plural, depending on how you mean them. If you are referring to a group as a unit, use a singular verb. If you are referring to the individuals within a group, use a plural verb.

- In terms of licenses, a **dozen is** probably enough for your company.
- In terms of users, a **dozen have** already registered.

Examples

- The employees are in the building.
- The employees is in the building.
- Everyone has chosen a desktop background.
- Everyone have chosen a desktop background.
- There are two ways to fix this problem.
- There is two ways to fix this problem.
- The news from the sales department is positive.
- The news from the sales department are positive.

than_then_confusion

Short Description

The words *than* and *then* can cause confusion.

Use **than** in comparison statements.

- This processor is faster than that one.
- The rewards package has more benefits than the original package.

Use **then** to refer to a time other than now.

- First turn on the computer, and then open a program.
- Back then, touch screens hardly existed.

Examples

- The hotel costs more than the hostel.
- The hotel costs more then the hostel.
- The shipment arrives then.
- The shipment arrives than.

there_their_confusion

Short Description

The words *there*, *their*, and *they're* can cause confusion.

Use **there** to refer to a location.

- The company has two branch offices there.
- We have started a new line of products over there.

Use **their** to mean *belonging to them*.

- The company has contacts at their headquarters.
- We have started to make inroads in their market.

Use **they're** as a contraction of *they are*.

- The customers have indicated that they're ready for something new.
- They're impressed by that presentation.

Examples

- They're not well.
- Their not well.
- Their package arrives tomorrow.
- There package arrives tomorrow.
- There is a meeting.
- They're is a meeting.

to_too_confusion

Short Description

The words *too* and *to* can cause confusion.

Use **too** only when you mean *also* or *in excess*.

- When the antivirus sweeps your hard drive, it will sweep peripheral devices too.
- If you put too much oil in the tank, it will overflow.

Use **to** for all other senses of the word.

- Press here to start.
- Clicking here will lead you to the home page.

Examples

- In order to check the system, the administrator must press the button.
- In order too check the system, the administrator must press the button.
- The administrator works too much.
- The administrator works to much.
- Me too.
- Me to.

uncountable_nouns

Short Description

Some nouns are uncountable and:

- do not have plural forms
- are used with singular verbs
- are never preceded by *a* or *an*

Examples of uncountable nouns are the following:

aluminum	applause
biology	blood
boating	butter
dancing	dust
equipment	flour
furniture	gold
harm	hockey
homework	ice
information	leather
luggage	meat
metainformation	misinformation
money	numerics
oxygen	photography
porcelain	publicity
research	rice
smoking	soccer
steel	sunshine
transportation	water
weather	wool

Examples

- Information is available at the company website.
- An information is available at the company website.
- New trainees are sometimes afraid to ask for advice.
- New trainees are sometimes afraid to ask for advices.
- I have ordered a piece of meat.
- I have ordered a meat.
- I have ordered a piece of meat.
- The students give an applause to the performer.

- We sell gold.
- We sell a gold.

unknown_word

Short Description

Check the spelling of this word.

Examples

- word
- wod
- example
- exmpl

use_comma_after_introduutory_phrase

Short Description

Do not overuse commas. Break long sentences with commas into shorter sentences to make the sentences clearer.

However, you can use commas after introductory (infinitive) phrases or adverbial clauses. An introductory phrase usually starts with a preposition.

Examples

- During the blending process, you can decrease the speed up to three times by pressing the button.
- During the blending process you can decrease the speed up to three times by pressing the button.
- With Unity Pro, you can use many programming languages.
- With Unity Pro you can use many programming languages.
- Despite the new product, the company makes no profit.
- Despite the new product the company makes no profit.

use_comma_after_subordinate_phrase

Short Description

Add a comma between the end of a subordinate clause and the beginning of the main clause. The main clause is likely to begin with a noun or a command (imperative verb).

Subordinate clauses are introduced by words such as

- although
- as

- because
- before
- but
- if
- once
- since
- than
- that
- though
- unless
- until
- when

Examples

- To learn more about this security update, go to the company website.
- To learn more about this security update go to the company website.
- Although the product is new to the division, it is already receiving good reviews.
- Although the product is new to the division it is already receiving good reviews.

use_comma_with_parenthetical_expressions

Short Description

Use commas before and after the following expressions, unless other sentence punctuation is called for:

and so on

or the like

and so forth

Examples

- The Bindings and Profiles specifications, and so on, give guidelines.
- The Bindings and Profiles specifications and so on give guidelines.
- It is possible to have multiple tags, or the like, on one object.
- It is possible to have multiple tags or the like on one object.
- She told me everything about her new department head, and so forth.
- She told me everything about her new department head and so forth.

use_end_of_sentence_punctuation

Short Description

Use a semicolon between independent clauses that are connected by a conjunctive adverb or a transitional phrase. Conjunctive adverbs include "also", "besides", "consequently", "conversely", "finally", "however", "instead", "likewise", "next", "otherwise", "similarly", "specifically", "still", "subsequently", "then", and "therefore". Transitional phrases include "at the same time", "for example", "in addition", and "on the contrary".

Examples

- AS/400e is sold worldwide; however, the US market remains a major focus.
- AS/400e is sold worldwide, however, the US market remains a major focus.

use_preposition_with_ing_verb

Short Description

Prepositions (e.g. *about, against, before, by, for, from, in, of, on, with*) are generally followed by *verb + -ing*.

Examples

- You can change the volume by adjusting the volume controls.
- You can change the volume by adjust the volume controls.
- Take these precautions before installing the software.
- Take these precautions before install the software.
- If you participate in doing anything, try to do it well.
- If you participate in do anything, try to do it well.
- I am working on improving my spelling.
- I am working on improve my spelling.
- Higher prices will discourage customers from buying.
- Higher prices will discourage customers from buy.
- Stop talking and get on with doing this exercise.
- Stop talking and get on with do this exercise.

use_verb_with_ing_verb

Short Description

The following verbs take the structure *verb + verb + -ing* when followed by another verb:

admit	postpone
avoid	prevent
defer	recommend
enjoy	risk
imagine	suggest
keep	

The manual recommends postponing installation.

Examples

- They recommend acquiring the images as digital files.
- They recommend to acquire the images as digital files.
- The steering committee has postponed making a decision until next week.

- The steering committee has postponed to make a decision until next week.
- Avoid using tables to format text documents.
- Avoid to use tables to format text documents.
- I enjoy working with my new colleagues.
- I enjoy to work with my new colleagues.
- My boss suggested cutting the budget.
- My boss suggested to cut the budget.

use_verb_with_object_and_infinitive

Short Description

The following verbs take the structure *verb + object + verb* without *to*:

make

let

This product lets customers start the machine easily.

Examples

- Hot weather makes the administrator feel tired.
- Hot weather makes to feel tired.
- They let the porter carry the luggage.
- They let the porter to carry the luggage.

use_verb_with_object_and_to

Short Description

The following verbs take the structure *verb + object + to + verb*:

advise

order

allow

persuade

enable

remind

encourage

show

forbid

teach

force

tell

instruct

warn

invite

The manual instructs users to start the machine.

The following verbs can also take the structure *verb + verb + -ing* when followed by another verb:

advise

allow

enable

The manual advises restarting the machine.

Examples

- The software enables users to share documents, pictures and video.
- The software enables to share documents, pictures and video.
- Your new cellphone automatically reminds you to return phone calls.
- Your new cellphone automatically reminds to return phone calls.
- They instruct customers to transfer their assets.
- They instruct to transfer assets.
- The teacher encourages the students to study in Germany.
- The teacher encourages to study in Germany.
- This console allows players to communicate with each other.
- This console allows to communicate with each other.

use_verb_with_to_and_infinitive

Short Description

The following verbs take the structure *verb + to + verb* when followed by another verb:

afford	decide	plan
agree	expect	prepare
aim	fail	promise
appear	forbid	refuse
arrange	hope	seem
ask	intend	tend
attempt	learn	threaten
beg	manage	want
claim	offer	

We aim to please.

Examples

- Choose this option if you intend to import all data.
- Choose this option if you intend importing all data.
- Elections tend to favor the incumbents.
- Elections tend to favoring the incumbents.
- The CEO decided to start a new policy.
- The CEO decided start a new policy.

use_will

Short Description

If you use "will", the sentence will be easier to read.

Examples

- I **will** try again, without the folder protection.
- I'm **going to** try again, without the folder protection.

where_were_confusion

Short Description

The words *where*, *were*, and *we're* can cause confusion.

Use **where** to refer to a *location*.

- Where is it?
- It is where I put it.

Use **were** as the past tense plural form of the verb *to be*.

- They were ready, but not anymore.
- The computers were delivered yesterday.

Use **we're** as a contraction of *we are*.

- We're on schedule.
- We're in the office.

Examples

- **Where** are the files?
- **Were** are the files?
- **We're** not very optimistic.
- **Were** are not very optimistic.
- **They were** not sure about our new products.
- **They where** not sure about our new products.

write_words_together

Short Description

Some words must be written together. Examples of these words are the following:

- albeit

- another
- itself
- nevertheless
- newspaper
- somewhat
- whereas
- without

The negated form of the verb *can* must be written as one word, *cannot*.

Note: In colloquial speech, some constructions can seem to contradict this rule. For example:

You can attend, or you can not attend if you choose.

In this case, however, *not* is negating the verb *attend*, not *can*. Writing *can not* separately is therefore correct. Constructions like this one are easy to understand when spoken but can be ambiguous when written. It is therefore best to rephrase them.

Examples

- Select another file.
- Select an other file.
- The program closes itself automatically after every update.
- The program closes it self automatically after every update.
- The supplier cannot come.
- The supplier can not come.
- The supplier does not need to come.
- The supplier can not come.

wrong_comparative_or_superlative

Short Description

General Guidelines

Adjectives that consist of two syllables and that end in *-y* form the comparative and superlative with *-ier* and *-iest*.

- easy, easier, easiest
- happy, happier, happiest
- trendy, trendier, trendiest

Exceptions

The comparative and superlative of some adjectives ending in *-ly* can also be formed with *more* or *most*.

- friendly, friendlier, friendliest
- likely, likelier, likeliest
- friendly, more friendly, most friendly

- likely, more likely, most likely

Examples

- This task is easier to do.
- This task is more easy to do.
- Customers prefer the older version.
- Customers prefer the more old version.
- The new product is worse than the previous one.
- The new product is more bad than the previous one.
- Acrolinx is the best grammar checking software.
- Acrolinx is the most good grammar checking software.

wrong_coordination

Short Description

The coordination seems to be incorrect. The adjective phrases on either side of the words *and* or *or* should be similar.

Examples

- The original or the old paper is changed.
- Original or the old paper is changed.
- The previous and the new versions are both popular.
- Previous and the new versions are both popular.
- I like both, white and silver.
- I like both, white and the silver.

wrong_sequence_of_words

Short Description

The sequence of words seems to be incorrect. A word might be missing, one of the marked words might be misspelled or unnecessary, or the word order might be incorrect.

Examples

- If this component is installed, check for possible mistakes.
- If is installed, check for possible mistakes.
- Your markings are stored in a GIF format.
- You markings are stored in a GIF format.
- The toolbar is in the fixed position at the top of the screen.
- The toolbar is the in fixed position at the top of the screen.
- This component adds customer-specific data to the resources.
- This component adds customer-specific data the to resources.

wrong_word

Short Description

The word or combination of words seems to be incorrect. It is possibly a typing or spelling mistake.

Examples

- Thank you.
- Thanks you.
- Mine is bigger than yours.
- Mine is bigger than your's.
- We would like to take it away.
- We would like to take it a away.
- Regarding your application, we cannot give you any information at this stage.
- In regards to your application, we cannot give you any information at this stage.
- Place the indicator on the control board.
- Place the indicator on he control board.
- Choose one or more user profiles to begin.
- Choose on or more user profiles to begin.
- In order to cut the budget, we have to skip this plan.
- In order cut the budget, we have to skip this plan.

wrong_extra_word

Short Description

The extra word is incorrect.

Examples

- The way I do it.
- The way how I do it.
- The students discuss their topic.
- The students discuss about their topic.
- What about this problem?
- What's about this problem?
- The location is near the town.
- The location is near of the town.

wrong_phrasing

Short Description

Certain phrases are fixed in English:

- do not under any circumstances do something
- for some time
- for that reason
- how something looks *or* what something looks like

- in any case
- in case of emergency
- in detail
- in my opinion
- in other words
- in particular

Examples

- Do not under any circumstances open the door when the vehicle is in motion.
- Do under no circumstances open the door when the vehicle is in motion.
- The server administrator has been trying to restart the computer for some time.
- The server administrator has been trying to restart the computer for some times.
- For safety reasons, make sure that the engine is switched off before opening the hatch.
- Because of safety reasons, make sure that the engine is switched off before opening the hatch.
- Make sure that you tell the maintenance personnel what the damaged part looks like. Make sure that you tell the maintenance personnel how the damaged part looks.
- Make sure that you tell the maintenance personnel how the damaged part looks like.
- In case of emergency, go to the nearest exit.
- In emergency case, go to the nearest exit.
- Describe the error in detail.
- Describe the error in details.
- In my opinion, this product performs well.
- After my opinion, this product performs well.
- In other words, place the wheel in a locked position.
- With other words, place the wheel in a locked position.
- That component in particular is vital.
- That component in particularly is vital.

wrong_verb_form

Short Description

The verb form is incorrect. Make sure that there is not a mismatch of verb tenses.

can, cannot, should, have to + infinitive verb

have, be + past participle

Examples

- These outputs can be used for pre-stop.
- These outputs can used for pre-stop.
- The system can process multiple types of data.
- The system can processed multiple types of data.
- If you have not yet applied the change, first determine the size of the application server.
- If you do not have applied the change, first determine the size of the application server.
- They have to install the application.

- They have to installed the application.

your_confusion

Short Description

The words *your* and *you're* can cause confusion.

Use **your** to mean *belonging to you*.

- May I borrow your pen?
- Please review your task.

Use **you're** as a contraction of *you are*.

- You're available.
- You're the administrator.

Examples

- You're excellent at that game.
- Your excellent at that game.
- This meeting is about your proposal.
- This meeting is about you're proposal.
- I borrowed money from your wife.
- I borrowed money from you're wife.
- Your administrator let me look into your system.
- You're administrator let me look into you're system.