Cross-linguistic annotation of narrativity for English/French verb tense disambiguation

Cristina Grisot[*], **Thomas Meyer**[†]

[*]University of Geneva
Department of Linguistics
1211 Geneva, Switzerland
cristina.grisot@unige.ch

[†]Idiap Research Institute
Rue Marconi 19
1920 Martigny, Switzerland
thomas.meyer@idiap.ch

Abstract

This paper presents manual and automatic annotation experiments for a pragmatic verb tense feature (narrativity) in English/French parallel corpora. The feature is considered to play an important role for translating English Simple Past tense into French, where three different tenses are available. Whether the French *Passé Composé, Passé Simple* or *Imparfait* should be used is highly dependent on a longer-range context, in which either narrative events ordered in time or mere non-narrative state of affairs in the past are described. This longer-range context is usually not available to current machine translation (MT) systems, that are trained on parallel corpora. Annotating narrativity prior to translation is therefore likely to help current MT systems. Our experiments show that narrativity can be reliably identified with *kappa*-values of up to 0.91 in manual annotation and with F1 scores of up to 0.72 in automatic annotation.

Keywords: verb tense, annotation, machine translation

1. Introduction

Parallel corpora of two and more languages show an increased use in various cross-linguistic, translation and machine translation studies. If such corpora are enriched by linguistic annotation, they can provide even more insight in contrastive differences between languages, such as the distribution of referential expressions or the increased or decreased use of cohesion devices in original and/or translated texts.

In this paper, we make use of parallel corpora in order to analyze verb tense for the English/French (EN/FR) language pair. Manual annotation was performed in order to detect *narrative* and *non-narrative* usages of the EN Simple Past tense (SP). Narrativity is a binary pragmatic feature: in narrative usages, the SP expresses eventualities (events/states) that are ordered in time, while non-narrative usages express un-ordered state of affairs in the past. This distinction allows to disambiguate the usages of the EN SP when translating into FR. For this language pair there exist translation divergencies because the EN SP can be translated by up to three different FR tense forms, that depend on (non-)narrative context: Passé Simple (PS), the Passé Composé (PC) and the Imparfait (IMP).

We mention related work in Section 2, and describe the data used in Section 3. The manual annotation procedure is illustrated as well as a possible way to detect *narrativity* automatically in Sections 4 and 5 respectively. The annotated resources and the automatic disambiguation model will be available at `http://www.idiap.ch/project/comtis/resources`.

2. Related Work

Verb tense has frequently been associated to narrative contexts in various frameworks, such as in Discourse Representation Theory (DRT) (Kamp and Reyle, 1993), where it was argued that each verb tense encodes a certain context of temporal information. In Segmented Discourse Representation Theory (SDRT), Lascarides and Asher (1993) focused on discourse being structured in segments that are related through discourse relations (narrative, explanation, etc.). Smith (2003) discussed discourse modes based in textual structure and aspect. If these theories focused 'only' on linguistic information, we consider narrativity to consist of pragmatic information (explicit or implicit) and to represent a *cognitive* (as opposed to *logic* in (S)DRT) discourse relation (Hobbs, 1979; Mann and Thompson, 1988; Sanders et al., 1992), that is expressed lexically through verb tense and connectives (that are language-specific) and can occur in any type of stylistic register. Additionally we focus on contrastive analyses of the translation divergencies that exist with respect to narrativity for the EN/FR language pair.

3. Data

For the annotation experiments described below, we made use of the parallel corpus that has been provided by Grisot and Cartoni (2012). The authors studied the discrepancies between theoretical descriptions of verb tenses and their use in parallel corpora. The corpus consists of texts in EN and their translations in FR belonging to four different genres with the following distribution: literature, 18%, journalistic, 18%, parliamentary discussions, 31% and legislation, 33%.

In example sentence 1 below, there are two events, i.e. *the marriage that happened* and *the wealth which was added*. The second event is presented in relation to the first (first he got married and then he added to his wealth), which is why the SP verbs *happened* and *added* are in narrative usage. In the second example, there are three states (*was a single man, lived* and *had a companion*) that describe the owner of the estate. States are not temporally ordered, which is why this example illustrates the non-narrative usage of the SP.

> **1**: *By his own marriage, likewise, which **happened** soon afterwards, he **added** to his wealth.* (Literature Corpus: J. Austen, Sense and Sensibility)
>
> **2**: *The late owner of this estate **was** a single man, who **lived** to a very advanced age, and who for many years of his life, **had** a constant companion and housekeeper in his sister.* (Literature Corpus: J. Austen, Sense and Sensibility)

By examining such samples of the EN SP in these parallel corpora, we found that the most frequent verb tenses that render the semantic and pragmatic domain of the SP in FR are PS, PC and IMP. Grisot and Moeschler (2014) made the hypothesis and confirmed through a pilot annotation experiment that the FR PS and PC have a *narrative usage* and the IMP has a *non-narrative usage*. However, these three tenses in FR are all possible translations of the EN SP, which is why the authors further hypothesized that semantic and pragmatic equivalences between the usages of the SP in EN and the usages of the three tenses in FR translation can be established.

We argue in this paper that narrativity is a relevant disambiguation criterion for the EN SP in order to find the correct FR verb tense. Narrative usages of the SP, are translated into FR by either PC or PS (narrative eventualities are temporally/causally related, and discourse relations can be explicitated through the insertion of connectives such as *and then, because*). Non-narrative usages of the SP, are translated in FR by IMP (non-narrative eventualities are not temporally/causally related or occur simultaneously, and relations can be explicitated through insertion of connectives such as *and* or no connective at all).

4. Manual annotation of narrativity

A manual annotation experiment was conducted to empirically test if the narrative and non-narrative usages of the SP can reliably be detected in EN. Two EN native speakers went through a training phase in order to check whether the instructions given were clear. The annotators had to annotate 10 text excerpts with SP occurrences and explained orally their reasoning. The results of the training were then discussed together with the authors of this paper.

The annotation guidelines included: (a) a definition of narrativity (b) the explanation of each usage (narrative and non-narrative) as explained above, with two examples for each usage, (c) the instruction to read each excerpt, to identify the verb highlighted and to decide if in context, the highlighted verb is part of the underlying theme and the connective *and then* could be added without changing the meaning (the verb would have a narrative usage) or not (non-narrative usage).

The data used for the annotation experiment was taken from the parallel corpus presented above. From this corpus, a subset of 458 excerpts (that we call items that all contain occurrences of the SP) was given to the two human annotators. For each item, the sentence with the SP verb, as well as one sentence before and/or after have been provided for sufficient contextual information.

The results of the human annotation experiments have been analyzed in three steps. As a first step, it can be tested whether different raters produced consistently similar results, so that one can infer that the annotators have understood the guidelines and that there was no agreement just by chance. In our annotation experiment, the two annotators agreed on 325 items (71%) and disagreed on 133 items (29%). This results in a $kappa$ value of 0.42, which is above chance, but not high enough to point to entirely reliable annotations (values around 0.6-0.7).

Error analysis revealed that the main source of errors is the length of the temporal interval between two eventualities perceived differently by the two annotators (due to ambiguity between temporal sequence or simultaneity, each of them corresponding to narrative, respectively, non-narrative usage). This has been corrected in a second annotation round, where the insertion of a connective is expected to force a narrative or a non-narrative reading.

Disagreements were thus resolved in a second annotation round by two new annotators, on a clean corpus containing 439 items. Annotators have been asked to insert a discourse connective in order to explicitate the implicit relation existing between eventualities. The connectives *and then/before* signaling temporal sequencing and *because/thus* for causal relations were proposed by annotators for the *narrative* label. For the *non-narrative* label, the connective *and* expressing simultaneity or no connective possibly inserted have been proposed. Note that we did not alter the sentences to annotate by actually inserting a connective, this was just to guide the annotators' choice. Inferring connectives is only a technique to render explicit the existent but implicit causal or temporal relations.

The inter-annotator agreement here was 0.91, signaling very strong and reliable agreement. Only 4 items of disagreement were found and can be considered as outliers in the data and were discarded from the corpus that totals in 435 items.

The data consisting of the annotators' agreements from both rounds has also been used for evaluation of the EN SP against the tenses used in the target language FR, taken from the parallel corpus. The narrative usages identified by annotators correspond to translations by PS/PC and the non-narrative usages correspond to translations by IMP in 80% of the cases. This leaves 20% of the corpus where annotators agreed on the narrativity label but where there is no consistency with the tense used in FR. We plan to look more closely at these 20% in future work.

5. Automatically labeling narrativity

The 435 correctly annotated instances of narrativity (257 narrative, 178 non-narrative), after resolving the disagreements as described in Section 4, have been used entirely for training a Maximum Entropy (MaxEnt) classifier with the Stanford Classifier package (Manning and Klein, 2003). Testing was performed on a smaller and earlier manually annotated sub-portion of the corpus with the same genre distribution, consisting of 118 labeled verbs: 75 instances of narrative and 43 of non-narrative.

From the training and test sets we extracted the features described below from syntactic parse trees obtained by pars-

ing the EN data with Charniak and Johnson's constituent parser (2005)). Furthermore, a TimeML parser (Verhagen and Pustejovsky, 2008) is used for features of temporal ordering of events in the sentences.

paragraphVerb word form. The English word form of the verb to classify as it appears in the text.

Neighboring verb word forms. We not only extract the verb form to label, but also all other verbs in the current sentence, thus building a 'bag-of-verbs". The value of this feature is a chain of verb word forms as they appear in the sentence.

Position. The numeric word index position of the verb in the sentence.

POS tags. From dependency parsing, POS tags for all words in the sentence are generated and output by the parser. As a separate feature, we concatenate the POS tags of the occurring verbs, i.e. all POS tags such as VB, VBN, VBG, etc., as they appear after the parsing.

Syntax. Similarly to POS tags, we get the syntactical categories and tree structures for the sentences from the Charniak parser.

Temporal markers. With a hand-made list of 66 temporal discourse markers we detect whether such markers are present in the sentence and use them as bag-of-word features. These e.g. consist of *while, since, weeks/days after, before, subsequently, repeatedly* etc.

Type of temporal markers. In addition to the actual marker word forms, we also consider whether a marker rather signals synchrony or asynchrony, or may signal both (e.g. *meanwhile*).

Temporal ordering. The TimeML annotation language tags events and their temporal order (FUTURE, INFINITIVE, PAST, PASTPART, etc.) as well as verbal aspect (PROGRESSIVE, PERFECTIVE, etc.). We thus use these tags obtained automatically from the output of the Tarsqi toolkit.

With these features, the MaxEnt classifier performs at 0.72 F1 score (weighted mean of precision and recall). Out of the 118 test instances, the classifier correctly annotates 90 items which corresponds to 76.27%. As a baseline to compare against, the majority class in the test set (narrative) would account for only 64% of correctly classified instances. Moreover, also the $kappa$ value for inter-class agreement is 0.46 with the classifier and is even a bit higher than the one obtained in the first manual annotation experiment. The classifier therefore is stable enough to automatically label the SP verbs in the EN side of a large parallel corpus that can then be used for detailed translational studies and/or for training machine translation systems. For the latter, it has been shown that statistical machine translation systems based on the narrativity feature can improve translation quality by up to +0.2 BLEU points and in up to 10% of all cases of translation for the EN SP (Meyer et al., 2013).

6. Conclusion

Based on a contrastive English/French corpus analysis, we identified the pragmatic feature of narrativity that disambiguates the usages of the English Simple Past tense, which then helps to find the corresponding past tense in French. We showed that after two annotation experiments and by using linguistic tests, the inter-annotator agreement is reliable and narrativity can be identified as a feature that correlates well with the tense usage in French. Further evidence was found by building an automatic classifier that labels narrative and non-narrative instances at the human performance level.

In future work we will refine the narrativity feature and examine its relation to other tense-specific semantic features such as boundedness, which in turn could also help to improve the classifier's performance via additional features. Moreover, as the classifier can label a large amount of data, we will integrate the labels output by the latter into the translation training processes of translation systems and evaluate these in terms of whether English Simple Past is more accurately translated in French.

7. References

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 173–180, Ann Arbor, MI.

Cristina Grisot and Bruno Cartoni. 2012. Une description bilingue des temps verbaux: étude contrastive en corpus. *Nouveaux cahiers de linguistique française*, 30:101–117.

Cristina Grisot and Jacques Moeschler. 2014. How do empirical methods interact with theoretical pragmatics? The conceptual and procedural contents of the English Simple Past and its translation into French. In J. Romero-Trillo, editor, *The Yearbook of Corpus Linguistics and Pragmatics*. Springer Verlag, Berlin, Germany.

Jerry R. Hobbs. 1979. Coherence and coreference. *Cognitive Science*, (3):67–90.

Hans Kamp and Uwe Reyle. 1993. *Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic Publishers.

Alex Lascarides and Nicolas Asher. 1993. Temporal interpretation, discourse relations and commonsense entailment. 16(5):437–493.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: towards a functional theory of text organization. *Text*, 8(3):243–281.

Christopher Manning and Dan Klein. 2003. Optimization, MaxEnt models, and conditional estimation without magic. In *Tutorial at HLT-NAACL and 41st ACL conferences*, Edmonton, Canada and Sapporo, Japan.

Thomas Meyer, Cristina Grisot, and Andrei Popescu-Belis. 2013. Detecting Narrativity to Improve English to

French Translation of Simple Past Verbs. In *Proceedings of the 1st DiscoMT Workshop at ACL 2013 (51st Annual Meeting of the Association for Computational Linguistics)*, pages 33–42, Sofia, Bulgaria.

Ted J. M. Sanders, Wilbert P. M. Spooren, and Leo G. M. Noordman. 1992. Toward a taxonomy of coherence relations. *Discourse Processes*, 15:1–35.

Carlota S. Smith. 2003. *Modes of Discourse: The Local Structure of Texts*. Cambridge University Press, UK.

Marc Verhagen and James Pustejovsky. 2008. Temporal processing with the TARSQI toolkit. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING), Companion volume: Demonstrations*, pages 189–192, Manchester, UK.