Temporal reference: empirical and theoretical perspectives. Converging evidence from English and Romance

PhD Dissertation defended on the 2nd of November 2015 by

Cristina GRISOT

University of Geneva

Summary

This thesis investigates the expression of temporal reference – the location of eventualities (states and events) in time – in natural languages by looking at four languages: English, French, Italian and Romanian. It is suggested that temporal reference is determined at the discursive level through a large variety of linguistic and pragmatic means. Among the linguistic means there are the grammatical categories of Tense and Aspect (generally referred to through the generic notion verbal tense), inherent temporal features of the verb phrase known as lexical aspect, also called Aktionsart, temporal adverbials such as ‘yesterday’, ‘last year’ and ‘in 1984’, as well as pragmatic temporal connectives such as ‘before’, ‘then’ and ‘at the same time’. The current state of the art on temporal reference lacks of a comprehensive cross-linguistically valid model that holds for typologically different languages, which goes beyond language specificities and which permits consistent contrastive analyses. The present study aims at filling in this gap by proposing a model of temporal reference situated in-between Theoretical and Applied Linguistics, Corpus Linguistics, Contrastive Analysis, Natural Language Processing and Machine Translation fields.

The research presented in this thesis is framed by two multidisciplinary projects aiming at improving the results of statistical machine translation systems in terms of textual coherence by providing them semantic and pragmatic information. The interest in temporal reference in this interdisciplinary perspective originates in the apparently simple empirical question: how can one improve the translation of verbal tenses by automatic translation systems? In order to answer this empirical question, other related issues were addressed, such as which verbal tenses are problematic, what is the meaning of a verbal tense, how do humans process this temporal information in order to understand and translate it and does the translation of a tense depend on other temporal devices in the discourse? This research examines these questions, on the one hand, through revisiting the most relevant existent syntactic, semantic and pragmatic theories, as well as neurolinguistic and computational investigations of the principal ingredients of temporal reference in discourse, i.e. Tense, Aspect and Aktionsart, and on the other hand, through a rich and solid empirical research on multilingual translation corpora and experimental data.

From a theoretical perspective, the research is carried out in the relevance theoretic framework (Sperber & Wilson 1986; Wilson & Sperber 2012) according to which the linguistic expressions that a speaker utters, hence temporal reference and its principal ingredients as well, underdetermine the content that she communicates not only at the level of implicatures but also the propositional contents she communicates explicitly. The hearer must therefore recover inferentially the speaker’s intended meaning, at the
level of explicatures and of implicatures. Moreover, linguistic expressions encode conceptual and procedural types of information, which play an essential role for working out contextually the meaning of a sentence and of its constituents. This interpretative process is guided by the expectation of relevance and the quest for cognitive effects. The model proposed in this thesis for temporal reference makes use of this distinction.

From an empirical perspective, several types of methodology are used. Firstly, translation corpora containing natural instances of human communication are explored in order to investigate the usages and the translation of verbal tenses, as well as to identify translation divergences. Secondly, corpus data was investigated using the methodology proposed in the Contrastive Analysis field, according to which systematic language comparison must be carried out in three steps: monolingual description followed by the actual comparison based on a cross-linguistic valid tertium comparationis. Finally, offline experiments with linguistic judgment task made by two native speaker judges were carried out in order to test theoretical hypotheses suggested in the literature, as well as novel working hypotheses. Inter-judge agreement rates were assessed with the Kappa coefficient, a statistic measure which takes into consideration agreement due to chance. Furthermore, two novel techniques making use of the properties of translation corpora were used for fine-grained investigations, precisely translation spotting and cross-linguistic transfer of properties. The model defended in this thesis was developed based on empirical work carried out within both data-based (i.e. testing of theoretical predictions built based on the literature and data-driven (i.e. novel theoretical insights emerged from the analysis of the data) approaches.

The most significant findings of this research can be summarized as follows.

Firstly, temporal reference was defined narrowly and broadly. Narrow temporal reference represents the localization of an eventuality with respect to the moment of speech (before, at the moment of speech and after the moment of speech). Broad temporal reference represents the localization of an eventuality with respect to another eventuality (phenomenon classically treated at temporal sequencing). In order to have an accurate view of temporal reference in natural languages, the currently generic term verbal tense was decomposed into its constituent categories (Tense and Aspect applied to Aktionsart) and the role played by the category of Tense is deemphasized.

Secondly, it was shown with the example of the English Simple Past and its translation into French, Italian and Romanian that Tense encodes both conceptual and procedural information, precisely, a pro-concept TIME and procedural information operationalized in the experimental work as the [±narrativity] feature. Aspect expressing the speaker’s viewpoint represents procedural information and it was operationalized as the [±perfectivity] feature. Aktionsart expressing temporal information inherent to situation types represents conceptual information and it was operationalized as the [±boundedness] feature. Procedural information constrains the formulation of hypotheses about the explicit and the implicit content of an utterance whereas conceptual information contributes to the truth-conditional content of an utterance. Approaching the ingredients of broad and narrow temporal reference in these terms based on experimental work is original and represents a contribution to the general state of the art.

Thirdly, it is suggested that a verbal tense has contextual usages, associated with the various combinations of conceptual and procedural contents of the subsequent categories Tense and Aspect applied to Aktionsart. Cross-linguistically, there is little variation with respect to the conceptual information encoded by these categories. As
for procedural information, it triggers important cross-linguistic variation. From a methodological point of view, two quantitative measures are proposed for conceptual and procedural types of encoded information: inter-judge agreement rate measured with the Kappa coefficient and cross-linguistic variation measured in translation corpora. Moreover, this research provides a cross-linguistically valid tertium comparationis for verbal tenses, as well as a reanalysis of individual English, French, Italian and Romanian verbal tenses expressing past time.

Fourthly, guided by the need to have cognitive temporal coherence, the hearer treats information coming from Tense, Aspect and Aktionsart and their interrelations in a coherent manner. Temporal coherence takes places at the cognitive level with respect to the hearer’s processing of the utterance and, the discursive level, with respect to temporal discourse relations, their linguistic marking and inferential nature. Multifactorial analyses of the data have shown that the features [±narrativity], [±boundedness] and [±perfectivity] are correlated and that the values of one feature can be predicted based on the values of the other features. Additionally, a statistical generalized mixed model showed that they are statistically significant factors for predicting the verbal tense used in a target language.

Fifthly, it is suggested that establishing temporal reference is a cognitive principle and that all humans apply it regardless of their language. Consequently, languages are typologically classified according to which are the predominant linguistic means at their disposal: tensed-prominent, such as the languages investigated in this research aspect-prominent, such as Slavic languages, and tenseless languages, such as the Mandarin Chinese, in which aspectual information and pragmatic inferential means are fundamental for expressing temporal reference. The model defended in this research permits its development in further work for aspect-prominent and tenseless languages, and therefore, it makes predictions for natural language, which is cognitive faculty specific to the human species.

Finally, describing temporal reference in terms of its ingredients and operationalizing their encoded conceptual and procedural types of information as medium-coarse grained features, i.e. the [±narrativity] and [±boundedness] features, proved to be the appropriate approach in order to answer the empirical research question regarding the automatic translation of verbal tenses. More precisely, the features proposed were not only successfully implemented for automatic processing but also their implementation in Natural Language Processing and their application to Statistical Machine Translation produced significant improvements of the results of these systems. These ameliorations represent in the same time an empirical indirect but solid validation of the theoretical model defended in this thesis.

To conclude, this research provides solid empirical findings for basic hypotheses defended in Relevance Theory and it proposes novel theoretical insights, it is a plea in favour of the necessity to have a rich empirical basis for theoretical linguistic models using complementary methods, such as corpus and experimental investigations, and it suggests an innovative comprehension of temporal reference in natural languages.