How empirical methods interact with theoretical pragmatics?1

The conceptual and procedural contents of the English Simple Past and its translation in French

Cristina Grisot and Jacques Moeschler2

One major theoretical issue that has dominated the field of pragmatics for the last twenty years is the conceptual vs. procedural distinction and its application for verb tenses. In this paper, we address this distinction from both theoretical and empirical perspectives following a multifaceted methodology (work on parallel corpus, contrastive analysis methodology, offline experimentation and natural language processing application) and we place our study under the cover of empirical pragmatics.

1 Introduction

In this paper, we address the relation between empirical methods and theoretical pragmatics. In the last few years, linguists became aware of the numerous advantages of the collaboration between theoretical linguistics, more specifically theoretical pragmatics, and empirical linguistics, in particular, empirical pragmatics. In our view, empirical pragmatics investigates language use from both descriptive-theoretical and empirical perspectives. The empirical means considered in this study are corpora and experimental methods. We argue that theoretical pragmatics and empirical pragmatics need to join forces in order to provide more and more insight on the use of language. This is also applicable for corpora and experimental methods themselves, where their findings are complementary and allow a better view on the linguistic phenomenon of interest.

Theoretical pragmatics can be defined in a broad sense as the study of language in use, and in a narrow sense, as the study of how linguistic properties and contextual factors interact for utterances interpretation (Noveck and Sperber 2004). In verbal communication, linguistic, linked to the content of sentences (phonological, syntactic, semantic assigned by the grammar of each language) and non-linguistic properties (linked to them being uttered in a given situation, at a given moment by a speaker) of utterances are involved. One question pragmatics wants to answer is the exact role of each type of properties and their interaction. On one hand, Grice (1989) and neo-Gricean scholars (Gazdar 1979; Horn 1973, 1984, 1989, 1992, 2004, 2007; Levinson 1983, 2000) proposed an explanation based on conversation maxims and principles that guide conversation participants. On the other hand, relevance theorists (Sperber and Wilson 1986/1995, Blakemore 1987, 2002; Carston 2002; Moeschler 1989; Reboul 1992; Reboul & Moeschler 1994, 1998a) speak about a unique expectation of relevance that hearers have while participating in an act of communication. According

to relevance theorists, this expectation of relevance is sufficient for recovering the speaker’s meaning.

Theoretical pragmatics (both neo-Griceans, Relevance theorists as well as other pragmaticians) is thus concerned with phenomena related to the interpretation of utterances, including both explicit (thus in close relation to semantics) and implicit meaning. The main assumption is that propositional structures are systematically underdetermined and must be contextually enriched. Of great interest for the present study is the theoretical distinction between conceptual vs. procedural meaning, proposed by Blakemore (1987) within the framework of Relevance Theory. As Escandell-Vidal, Leonetti, Ahern (2011) argue, the conceptual/procedural distinction was first meant as a solution for the semantics/pragmatics division of labour and it has remained an important explanation for the contribution of linguistic meaning to utterance interpretation. A speaker is not expected to render more difficult than necessary his/her addressee’s task in obtaining a relevant interpretation. In this way, procedural meanings are encoded instructions that specify paths to follow during the interpretation process (manipulation of conceptual representations), more specifically to access the most relevant context. Wilson and Sperber (1993) attach cognitive foundations to the conceptual/procedural distinction and they mention a property for procedural meanings: conceptual representations can be brought to consciousness while procedures cannot. We are particularly interested in this distinction because of its highly debated application for verb tenses (Smith 1990; Wilson and Sperber 1993; Moeschler et al. 1994; Moeschler 2000, 2002, Saussure 2003, 2010; Amérons-Pons 2011; Moeschler et al. 1998, 2012; Grisot et al. submitted).

The two aims of this paper are (1) to illustrate our investigation of the conceptual and procedural meanings of verb tenses and (2) to argue for the benefits of combining two empirical methods, corpus analysis and linguistic experiments. In our study, we combined data from parallel corpora that served as stimulus composition for offline experiments (linguistic judgement task). Parallel corpora revealed variation in translation possibilities of a verb tense from a source language to a target language. Based on semantic and pragmatic theories we made hypotheses about the source of this variation and possible disambiguation criteria. Offline experiments allowed us to validate one of these criteria, as well as to propose new theoretic descriptions of the meaning and usages of verb tenses. We place this study under the empirical pragmatics cover.

Empirical pragmatics draws on theoretical pragmatics and corpus linguistics, adopting in the same time experimental methods. Probably the reason of existence of empirical pragmatics is the need of having consistent data for supporting or challenging current pragmatics theories, as well as proposing new models for the interpretation of linguistic phenomena. Of course, theoretical pragmatics makes use of data (built examples) which represents mainly the researchers’ own intuitions. This type of data has been criticized over the years mainly for its subjectivity and lack of replicability and this is why more robust (objective, quantifiable, replicable) data must be adopted, such as data from corpora and experiments. From the two types of experiments used in psycholinguistics and experimental pragmatics, only offline experimentation can be adopted more easily by empirical pragmatics because of the lack of material required (no necessity of a laboratory with electroencephalography EEG material\(^3\) or eye-trackers).

\(^3\) EEG is a procedure that measures electrical activity of the brain over time using electrodes placed on the scalp and it reflects thousands of simultaneously ongoing brain processes. Eye tracking is the...
There is one branch of pragmatics that has integrated experimental methodologies for testing pragmatic theories: experimental pragmatics. While theoretical pragmatics is rooted in philosophy of language and in linguistics, experimental pragmatics, drawing on pragmatics, psycholinguistics and psychology of reasoning, has taken over and reinterpreted the psycholinguistics’ sophisticated experimental methods (Meibauer and Steinbach 2011). For instance, Katsos and Cummins (2010) emphasize the relation between pragmatic theory and psycholinguistics’ experimental design: linguists benefit from experimental data confirming the psychological validity of their observations and provide critical evidence for cases that go beyond the reach of intuitive reflection, while psychologists benefit from a wide range of phenomena to study and of multiple theories provided by semantics and pragmatics. Recent experimental pragmatics (such as papers from the volume edited by Noveck and Sperber in 2004) has focused on phenomena such as indirect speech acts, metaphors, implicature, presupposition and, more generally, speaker meaning.

Finally, we would like to argue that empirical pragmatics has built a bridge to the Natural Language Processing (NLP) domain thanks to the robust type of data used. The NLP domain needs models of language interpretation inspired from theoretical pragmatics that can be adapted to machines. NLP also requires large amounts of data that allow quantitative analyses, statistical models and data for training linguistic software. Empirical pragmatics is able to provide NLP both linguistic models and empirical data.

This chapter is structured as follows: in section 2, we introduce from a general point of view the role and type of data in linguistics, and more particularly in semantics and pragmatics, as well as their advantages and limits; in section 3, we describe our case study by pointing out theoretical matters about verb tenses, our hypotheses, our empirical study on parallel corpus and offline experiments. We conclude our paper in section 4 by addressing the impact of the results of our experiment on theoretical matters about verb tenses and the importance of giving multiple sources of data for empirical pragmatics studies.

2 Type and role of data in empirical pragmatics

Nowadays, one can observe the increasing aspirations of linguists to have more robust and objective findings next to intuitive and subjective acceptability judgements or built examples. McEnery and Wilson (2001) underline that, broadly speaking, linguists have tended to favour the use of either introspective data (that is, language data constructed by linguists) or naturally occurring data (that is, examples of actual language usage). Nowadays, most linguists see these two types of data as complementary approaches, and not exclusive ones. Gibbs and Matlock (1999:263) and Gries (2002) argue that, although intuition may be poor as methodology for investigating mental representations, linguists’ intuitions are useful in the formulation of ‘testable hypotheses about linguistic structure and behaviour’.

Kepser and Reis (2005) point out that if introspective and corpus data have been the two main sources of data for theoretical linguistics until the mid-1990s, beyond that moment other sources have been considered, such as experimental (offline and online experiments), language acquisition, language pathologies, neurolinguistic, etc. They argue that linguistic evidence coming from different domains of data shed

process of measuring either the point of gaze or the motion of an eye relative to the head and it is used to investigate human thought processes.
diverse types of light on issues investigated having either similar results, thus validating the theory, or contradictory results, therefore opening new perspectives.

In what concerns natural occurring data, Table 2-1 provides an overview of kinds of linguistic data (from Gilquin and Gries 2009:5) presented in descending order of naturalness of production and collection (only corpora with written examples are produced for other aims than the specific purpose of linguistic research, thus the most natural).

Table 2-1: Kinds of linguistic data (sorted according to naturalness of production/collection)

<table>
<thead>
<tr>
<th>Data source</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1   corpora with written texts (e.g. newspapers, weblogs)</td>
<td></td>
</tr>
<tr>
<td>2   example collections</td>
<td></td>
</tr>
<tr>
<td>3   corpora of recorded spoken language in societies/communities where note-taking/recording is not particularly spectacular/invasive</td>
<td></td>
</tr>
<tr>
<td>4   corpora with recorded spoken language from fieldwork in societies/communities where note-taking/recording is spectacular/invasive</td>
<td></td>
</tr>
<tr>
<td>5   data from interviews (e.g. sociolinguistic interviews)</td>
<td></td>
</tr>
<tr>
<td>6   experimentation requiring subjects to do something with language they usually do anyway (e.g. sentence production as in answering questions in studies on priming or picture description in studies on information structure)</td>
<td></td>
</tr>
<tr>
<td>7   elicited data from fieldwork (e.g. response to “how do you say X in your language?”)</td>
<td></td>
</tr>
<tr>
<td>8   experimentation requiring subjects to do something with language they usually do, *on units they usually interact with (e.g. sentence sorting, measurements of reaction times in lexical decision tasks, word associations)</td>
<td></td>
</tr>
<tr>
<td>9   experimentation requiring subjects to do something with language they usually do not do, *on units they usually interact with, involving typical linguistic output (e.g. measurements of event-related potentials evoked by viewing pictures, eye-movement during reading idioms, acceptability/grammaticality judgements) *on units they usually do not interact with, involving production of linguistic output (e.g. phoneme monitoring, ultrasound tongue-position videos)</td>
<td></td>
</tr>
</tbody>
</table>

In this paper, we are interested in the first and the last type of data, namely corpora with written texts and data coming from experimentation where subjects are required to do something with language they do not usually do (using units they usually interact with involving typical linguistic output). We argue that both types of data are complementary and necessary in pragmatics research, and may be used within various frameworks of linguistic description and analysis.

Before presenting the advantages and difficulties, as well as the complementarity of both empirical methods used in this study, we will define briefly corpora and offline experiments and provide some important features.

2.1 Corpora

The well-known description of a corpus as being ‘a body of naturally occurring language’ (McEnery, Xiao and Tono 2006:4) is largely accepted in the Corpus Linguistics community, as well as other domain that work on corpora, such as Empirical Pragmatics or Translation Studies (Baker 1993, 1995). The same is true for corpora as having nowadays a machine-readable form, a feature that allows its compilation and analysis semi-automatically and automatically. In what concerns the size, corpora tend to be larger and larger thanks to the possibility to be tagged, compiled and analysed automatically, going as far as considering the World Wide Web as a whole a possible corpus. The most important aspect is having an appropriate match of the research goal and the corpus type and size.
Another feature of corpora is the number of languages and type of texts they contain, for example, monolingual or multilingual. Multilingual corpora can be of two main types: (a) parallel (or translation) corpora, containing source texts and their translation in one or several target languages, which can be unidirectional (from language A to language B) or bi/multidirectional, and (b) comparable corpora, containing non-translated or translated texts of the same genre. Each type can be used for specific research goals.

A first advantage of working on corpora is that they represent an empirical basis for researcher’s intuitions. Intuitions are the starting-point of any study but can be misleading and sometimes a few striking differences could lead to hazardous generalizations. Moreover, results of analyses of quantifiable data allow not only generalizations (through statistical significance tests) but also predictions through statistical analyses, such as correlations or multiple regression models, which are often used for investigating such a complex phenomenon as language.

Furthermore, multilingual corpora have quite naturally been used in contrastive studies. Contrastive Linguistics, also called Contrastive Analysis (CA), is “the systematic comparison of two or more languages, with the aim of describing their similarities and differences” (Johansson 2003:31) and it is often done by focusing one linguistic phenomenon. Mainly, the methodology used in a contrastive study consists of a first phase of monolingual description of the data (the phenomenon to be analysed), followed by the juxtaposition of the two or more monolingual descriptions and the analysis of the elements according to a tertium comparationis (James 1980, Krzeszowski 1990). We argue and develop later on in our case study that the necessary tertium comparationis for verb tenses should be defined in terms of cross-linguistic valid features, such as conceptual and procedural information.

The practice of contrastive languages comparison based on corpora has itself numerous advantages, specifically for new insights on the languages to be compared (which would have remained unnoticed in studies of monolingual corpora), the highlighting of language-specific features and the possibility of making semantic and pragmatic equivalences for the considered linguistic phenomenon between the source language (SL) and the target language (TL). In some cases, corpus-based studies with a contrastive perspective have applicable purposes, such as our case study that aims at modelling verb tenses for improving the quality of the texts translated by machine translation systems.

Another advantage is that data from corpora can be annotated (enriched) with semantic and pragmatic information, which allows more complex analyses. Annotation is “the practice of adding interpretative linguistic information to a corpus” (Leech 2005). Annotation is thus an enrichment of the original raw corpus. From this perspective, adding annotation to a corpus is providing additional value, which can be used for research by the individual or team that carried out the annotation, but which

---

4 Correlation is a monofactorial statistical method, which investigates the relation between one independent variable (the predictor) and one dependent variable (the phenomenon of interest). Correlation does not involve obligatorily causality between the two variables (they can be only associated) and explains only the linear aspects of the relationship (cf. Gries 2009, Baayen 2008).

5 Multiple regressions are multifactorial statistical methods, which investigate the relation between several independent variables (predictors) and one dependent variable, as well as their interactions. The relation between independent variables and the dependent variable can be linear or non-linear. (cf. Gries 2009, Baayen 2008).
can also be passed on to others who may find it useful for their own purposes. Several authors, such as McEnery (2003), Leech (2004) underline the fact that linguistic annotation increases the utility of a corpus. This happens by allowing both manual and automatic analysis and processing of the corpus and by assuring its multifunctional utilisation (the annotations themselves often reveal a whole range of uses which would not have been practicable unless the corpus had been annotated), as well as an objective record of analysis open to future analysis, decisions are more objective, reproducible, as well as insights to the corpus are more accessible to others and not only to the researcher himself. Due to automatic analysis of the corpus, annotated corpora are often used for training of NLP tools.

Corpus work is thus interesting when the researcher is concerned with a descriptive approach of the linguistic phenomenon considered, as well as the study of language as it is used (most of the time cotext and contextual information is also available in the corpus) in one language and cross-linguistically. Furthermore, corpus work allows the researcher to uncover on one hand, what is probable and typical and on the other hand, what is unusual about the phenomenon considered.

Corpus work has also some difficulties, such as the availability of multilingual corpora for less widespread languages or the predilection for ‘form-based research’ (interest in a specific grammatical form) as Granger (2003) points out. These difficulties constrain researchers to carry out their research manually, including building their corpus themselves (electronic version and alignment) and annotating it if they are interested in other phenomena than a specific grammatical form (such as semantic or syntactic categories). Another difficulty about corpus work is when the researcher is interested in less frequent phenomena (thus absent or with too few occurrences in corpus), in phenomena that are not lexically expressed therefore implicit (such as world knowledge used in inferences) and in the cognitive basis of language.

This is one reason why corpus data it is more and more combined with other types of evidence, such as experimentation. In what follows, we will briefly describe the use of experimentation in pragmatics and underline the complementarity between corpus work and experimentation.

### 2.2 Experimentation

In pragmatics, experimentation has been extremely useful for studying issues from the semantics/pragmatics interface and testing theories concerning the psychologically real competence native speakers have regarding semantics and pragmatics (Katsos and Breheny 2008).

One important distinction at the semantics/pragmatics interface has been proposed by Grice (1975/1989) between what is ‘said’ vs. what is ‘implicated’ within the entire meaning of an utterance. The first experimental study of the identification and labelling by ordinary speakers of what is ‘said’ vs. what is ‘implicated’ was Gibbs and Moise (1997). In their paper, Gibbs and Moise designed their experiments also to determine whether people viewed what is ‘said’ as being enriched pragmatically. They used five categories of sentences and participants had to choose between a

---

6 For example, Grivaz (2012, PhD dissertation) who studied causality in certain pairs of verbs in a very large corpus and with human annotation experiments, found that less frequent pairs had a good causal correlation while very frequent pairs had a small causal relation or even inexistent.

7 Cardinal (Jane has three children), possession (Robert broke a finger last night), scalar (Everyone went to Paris), time-distance (It will take us some time to get there) and temporal relations.
minimal vs. enriched interpretation. The temporal relation type of sentence as well as the two possible interpretations are given in example (1).

(1) ‘The old king died of heart attack and a republic was declared’.
   (a) **Minimal**: order of event unspecified
   (b) **Enriched**: the old kind died and then a republic was declared

The design of experiments consisted of manipulating the type of sentence, the instructions and the context. In the first experiment, the instructions consisted in explaining the two categories of interpretation of the sentence and no context was given. In the second experiment, the instructions consisted of more detail, including information about linguistic theories about the distinction between what is ‘said’ and what is ‘implicated’. In the last two experiments, contexts were provided in order to bias in one experiment for the implicature and in the other for the minimal interpretation.

Gibbs and Moise’s four experiments showed that speakers assume that enriched pragmatics plays a significant role in what is said: the enriched interpretation was preferred in the first 3 experiments but not in the last one where the context biased strongly for the minimal interpretation. Manipulation of instructions and training did not have any effect on the participants’ judgements.

In what concerns the fifth type of sentence from Gibbs and Moise’s experiments (and of interest in our case study), namely temporal relation between eventualities, we can underline that temporal sequencing (information provided through pragmatic enrichment that speakers are able to identify) is an inference drawn contextually\(^8\) (independent of the specific instructions that speakers received), which can be blocked by context biasing for the minimal interpretation (the unspecified order). In our case study, we will consider temporal sequencing as a feature in an annotation experiment and argue that it is a disambiguation criterion for usages of the English Simple Past.

Now we turn to experimentation as methodology used in empirical and experimental pragmatics and we point out two advantages of adopting it: (i) they make possible systematically control for confounding variables, and (ii) depending on the nature of the experiment, they permit the study of online processes (Gilquin and Gries 2009:9). One difficulty with experimentation is the artificial setting experiments require, as well as the complexity of the methodology to follow, including statistical methods used to analyze the data. If experimental pragmatics completely adopted the psycholinguistics methodology as well as the study of online processes (through EEG and eye-tracking tools), empirical pragmatics focuses mainly on offline experimentation, preserving the very essence of experimental studies: systematic manipulation of independent variables in order to determine their effect on dependent variables.

In what concerns the complementarity of the two empirical sources of data, Gilquin and Gries argue that corpus has a four-folded purpose in experimentation: (a) validator (the corpus serves as a validator of the experiment), (b) validatee (the corpus is validated by the experiment), (c) equal (corpus and experimental data are used on an equal footing) and (d) stimulus composition (the corpus serves as a database for the items used in experiments). They also note that corpus work deals with a larger range

\(^8\) In his Model of Directional Inferences (2000, 2002), Moeschler proposes the same prediction about temporal relations between eventualities. They have an inferential nature and are drawn based on contextual assumptions. They can be blocked (minimal interpretation) under certain specific linguistic and contextual conditions.
of data that can be investigated compared to experimentation, while experimentation allows the study of phenomena that are too infrequent on corpora. Corpora and experiments have thus advantages and disadvantages that are complementary and thus linguists tend nowadays to use both of these empirical methodologies.

Besides all this, we would add that data from experiments (namely from linguistic judgement task) are human annotated data and can be used for NLP as training for automatic classifiers.

This article considers data from experimentation (the 9th type of data in Gilquin and Gries’ classification), focusing on linguistic judgments made by subjects. If linguistic judgments were used mainly for acceptability/grammaticality tasks, nowadays (including in the experiments presented in this paper) linguistic judgement tasks concern all types of linguistic information. By presenting our case study, we aim to point out the complementarity of corpus work and experimentation for testing theoretic hypothesis, build description models and apply them to NLP.

In what follows, we provide a case study presenting our investigation on verb tenses and show how the methodology presented above has been used, as well as how the results of our study support our thesis about the advantages of combining corpora work and experimentation when doing empirical pragmatics research.

3 Case study

The case study presented in this article is incorporated within a research project9 that aims at improving the results of statistical machine translation (SMT) systems by modelling intersentential relations, such as those that depend on verb tenses and connectives. We investigated the ‘meaning’ of verb tenses, where the meaning is seen as consisting of both what is said and what is implicated. We deal thus with the semantics and pragmatics of verb tenses. Within the frame of empirical pragmatics, we studied verb tenses within Relevance Theory from a contrastive perspective (following specific methodology according to Contrastive Analyse domain) and based on parallel corpora and offline experimentation. Moreover, data from experimentation (human annotation) was used for automatic annotation and, furthermore, for training of a SMT system.

As Aménos-Pons (2011) correctly underlines, any approach to tenses must deal with the fact that they present a certain stability of some basic features, combined with a high adaptability at discourse level, that depends on contextual information (semantic and pragmatic) and world knowledge. A great challenge for linguists was, and remains, to know which of the features of verb tenses are stable and which are not.

Probably, one of the few generally accepted ideas about the meaning of verb tenses is the linguistic underdeterminacy thesis, as developed in Relevance Theory (Sperber and Wilson 1986/1995) and applied specifically to verb tenses by Smith (1990). According to it, verb tenses are defined as a referential category: they can be characterized as locating temporal reference for eventualities with respect to three coordinates (speech moment S, event moment E and reference point R as Reichenbach addressed them in 1947) through contextual enrichment following the expectation of optimal relevance. The consequence of this theory is that verb tenses

---

9 The COMTIS Project (Improving the Coherence of Machine Translation Output by modelling Intersentential Relations 2010-2013), a Sinergia interdisciplinary program funded by the Swiss National Science Foundation (no. CRSI22-127510).
do not have several meanings but several usages corresponding to different contextual interpretations.

In the literature, two main trends are opposed regarding the nature of the encoded content verb tenses: on one hand, verb tenses have only rigid procedural meanings help the hearer reconstruct the intended representation of eventualities (Améanos-Pons 2011). Saussure (2003) goes in the same direction and proposes algorithms to follow, consisting of the instructions encoded by verb tenses, in order to grasp the intended meaning of a verb tense at the discourse level.

On the other hand, verb tenses are seen as having both procedural and conceptual contents, as argued in Moeschler (2002) and Moeschler et al. (2012). In Moeschler et al. (2012), we take over an idea from Moeschler (2002) and argue that verb tenses have a stable semantic core and a flexible pragmatic set of properties adaptable in context usage. They argue for a stable semantic core consisting of different configurations of Reichenbachian coordinates (Reichenbach 1947) S, R and E distributed according to two possible relations simultaneity or anteriority between S and R, where the relation between E and S is inferred. S, R and E are variables that are to be saturated in each context, producing several contextual usages of each verb tense. The assumption is that the conceptual content of verb tenses (specifically, a specific configuration of temporal coordinates S, R and E) behaves like pro-concepts (Wilson 2011, Sperber and Wilson 1998: 15). Pro-concepts are semantically incomplete, they are conveyed in a given utterance and have to be contextually worked out. Once the enrichment process is completed, the propositional form of the utterance is also available. This temporal information is not defeasible, i.e. it can not be cancelled. Temporal coordinates S, R and E combine with the predicate’s lexical aspect, in order to allow the calculation of the aspectual class (state, process, event). This semantic configuration is the skeleton of the meaning of each verb tense, which is enriched with contextual information and world knowledge in the inferential interpretation process. This semantic skeleton is stable and robust and represents the conceptual meaning of verb tenses.

In what concerns the status of temporal coordinates, Saussure and Morency (2012) argue that tenses encode instructions on how the eventuality is to be represented by the hearer through the positions of temporal coordinates. They consider thus that temporal location with the help of S, R and E is of a procedural nature. We will show later on in this paper that experimental studies revealed the contrary: the configuration of temporal coordinates is of a conceptual nature, that they are variables that are saturated contextually.

The procedural content of verb tenses, on the other hand, includes instructions to look for three types of information: if R is part of a series of points of reference available in the context and thus, eventualities are temporally sequenced (procedure called [± narrative]), if there is a perspective or a point of view on the eventuality presented (procedure called [± subjective]) and if this perspective is explicit or implicit (procedure called [± explicit]). The experimental work that we conducted (see section 3.3.3) showed that the [± narrative] feature includes temporal sequencing (inferential temporal relation as in Gibbs and Moise's experiments described in section 2.2) and causal relations holding between eventualities (cf. Moeschler 2003, 2011 for the relation between causality and temporal sequencing).

Another important point in the model described in Moeschler et al. (2012) is that the specific combination of conceptual content - temporal coordinates (S, R and E) and procedural content - [± narrative], [± subjective] and [± explicit] characterises...
contextual usages of verb tenses and should be considered as the meaning of each verb tense. For this point Moeschler et al.’s analysis joins Aménos-Pons (2011) who assume that “tenses do not encode temporal relations; they come only as a by-product of the tense meaning in specific environments”.

In this paper we adopt the view proposed by Moeschler et al. (2012) and we bring new arguments, as well as evidence from experimental work, that support the procedural and conceptual nature of the information encoded by verb tenses expressing past time in FR and EN.

3.1 Our hypotheses

Parallel corpora consisting of several stylistic genres revealed the four most frequent translation divergences: (a) from English (EN) into French (FR): the Simple Past (SP), the Simple Present and the Present Perfect (PresPerf), and (b) from FR into EN: the Passé Composé and Présent. In a first research phase, we chose to investigate the translation of the EN Simple Past in FR, where its semantic and pragmatic domain is rendered through the Passé Simple (PS), Passé Composé (PC) and Imparfait (IMP). In order to grasp the meaning of the EN SP, we assume that the distinction between conceptual and procedural types of information is very important.

Our assumptions in this study are: (1) verb tenses encode conceptual and procedural information and (2) conceptual and procedural contents explain cross-linguistic variation. In what concerns the first hypothesis, we defend and bring evidence from offline experiments that procedural meanings of the English SP are inaccessible to consciousness and hard to describe in conceptual terms, while its conceptual meaning is accessible to conscious thinking and can be conceptualized. We also argue that the conceptual content of verb tenses (specifically, a specific configuration of temporal coordinates S, E and R) behaves more like pro-concepts in that they are conveyed in a given utterance and have to be contextually worked out.

In what concerns our second hypothesis, we assume that conceptual and procedural contents of verb tenses explain their cross-linguistic variation revealed by parallel corpora analysis. We suggest that verb tenses have several contextual usages. For each usage, there is a linguistic feature related to their conceptual and procedural contents, which is the most salient information for launching different processes of pragmatic enrichment according to the principle of relevance. Thus, a verb tense can have in the interpretation process several usages triggered either by its conceptual or its procedural contents and these usages are linked to the cross-linguistic variation observed in translation corpora.

3.2 Data from parallel corpora with a contrastive perspective

In Grisot and Cartoni (2012) we studied the discrepancies between theoretical descriptions of verb tenses and their use in parallel corpora. The corpus that we investigated consists of texts in EN and their translations in FR belonging to four different genres (literature, journalistic, legislation and EuroParl). The qualitative and quantitative analysis of the corpus was done in two steps: (a) a first monolingual step in order to see which tenses occur in the corpus and calculate their frequency in source language (SL) and (b) a second bilingual step in order to identify the tenses used as translation possibilities in a target language (TL) for a certain tense in source language, as well as calculate their frequency. Analysis of frequency of tenses in SL provided information about tenses that are possible candidates for being problematic for machine translation systems, with the assumption that frequent tenses, if wrongly
translated, decrease the quality of the translated text. Bilingual analysis with focus on identifying verb tenses used as translation possibilities in TL for ambiguous tenses in SL revealed that the Simple Past is translated into FR using mainly three tenses (PS, PC and IMP) and that the Present Perfect is translated using two tenses (PC and Présent). These are two of the translation divergences shown by analysis of parallel corpora.

The ambiguity of EN Simple Past, as well as the Present Perfect, is emphasized when we look at their translation in FR. In order to improve their translation by SMT systems, these tenses must be disambiguated. Following the CA’s methodology, the SP and the Present Perfect, as well as the FR tenses used for their translation, must be compared in three steps. The first step consists of the monolingual description, followed by bilingual juxtaposition of the two monolingual descriptions and finally, their analysis according to the tertium comparationis defined in terms of conceptual and procedural contents.

Now in what concerns the SP, known as preterit, it describes an action or state as having occurred or having existed at a past moment or during a past period of time that is definitely separated from the actual present moment of speaking or writing. Comrie (1985: 41) emphasized that the SP “only locates the event in the past, without saying anything about whether the situation continues up to the present or into the future”. Radden and Dirven (2007:219) argue that the use of the SP to express bounded past situations, presented as a series of events, typically in narratives, as in (2). The individual events from example (2) are temporally ordered (signalled by the coordination and the conjunction and) and are thus interpreted as being successive.

(2) I grabbed his arm and I twisted it up behind his back and when I let go his arm there was a knife on the table and he just picked it up and let me have it and I started bleeding like a pig. (Labov and Waletzky 1967, quoted in Radden and Dirven 2007: 219)

The verb tenses used in FR for translating the SP are, as we have already noted, the PC, PS and PC. Their monolingual description is generally done in terms of: (a) “tense with two faces” (Martin 1971) because of the possibility to express both past and present time for the PC, (b) expressing a past event completely accomplished in the past with no connection to present time (Grevisse 1980, Wagner and Pinchon 1962) and used in contexts where events are temporally ordered (Kamp and Rohrer 1983) for the PS, and (c) tense of background information (Weinrich 1973) for the IMP. The focus on the accomplishment of the event in the past is the feature that distinguishes the PS from the PC, the second one expressing a link to present time, while perfectivity is a feature that distinguishes the PS from the IMP, the former being perfective and the latter imperfective.

Given these monolingual descriptions, when juxtaposed, we can observe the multitude of facets for describing these four tenses: in terms of temporal location (time preceding, simultaneous or even following speech moment), grammatical aspect (perfective or imperfective), discursive grounding (foreground or background) and relation to other eventualities (temporally ordered or not). Another point that can be observed is the lack of one-to-one correspondence between the several meanings of the SP and the three FR tenses used for its translation. In Grisot, Cartoni and Moeschler (to appear), we argue that the meaning of these verb tenses should be cross-linguistically established in terms of their conceptual and procedural information, and more specifically that the procedural information [± narrativity] is a disambiguation criterion for the usages of the SP. In this study, we go in the same direction, showing that the [± narrativity] feature is indeed procedural (through experimental work presented in section 3.3.3), and that items of SP annotated by two
human annotators as having a narrative usage correspond in parallel corpora to translation through either PS or PC and items annotated as having a non-narrative usage correspond to translation through an IMP (detailed results provided in section 3.3.3).

In what concerns the EN Present Perfect, the grammatical combination of present tense and perfect aspect, is used to express a past eventuality that has present relevance. The same grammatical combination exists in other languages also, such as FR (known as the Passé Composé) but where it can also express eventualities accomplished in the past. In EN, there is a competition between the SP and the PresPerf for referring to past time eventualities, with the particularity that PresPerf is incompatible with adverbials expressing define past time. The first annotation experiment considered the competition between SP and PresPerf forms for expressing past time eventualities, showing that each verb tense has conceptual meaning and it can be easily dealt with by human annotators (section 3.3.2)

A benefit of parallel corpora is availability of context and cotext, information that facilitates establishing semantic and pragmatic equivalence for each verb tenses. This information is crucial as regards the meaning of verb tenses.

From the corpus described above, we used a subset of 30 excerpts randomly selected (that we call items and all contain occurrences of the SP or Present Perfect) for the first experiment and 458 items (containing occurrences of the SP) for the second experiment. In what follows, we describe and provide the results of annotation experiments.

### 3.3 Data from offline experiments

Experimental work we have conducted brought arguments for the hypothesis that verb tenses encode both conceptual and procedural information. Conceptual information concerns different combinations of Reichenbachian temporal coordinates, which are contextually saturated variables. Procedural information concerns instructions relating the reference point R of an eventuality to reference points of other eventualities from the cotext, in order to check their temporal order. In this section, we will provide the general framework of our experiments (participants, procedure and evaluation), followed by the presentation of the two experiments and their results.

#### 3.3.1 Design of experiments and participants

The two annotators were native speakers of EN with basic knowledge of FR. They were asked to follow the instructions (given below for each type of information annotated) and went through a training phase in order to check whether the instructions given were clear. For the effective annotation task, annotators received a file with the total number of excerpts that were taken from the EN part of the parallel corpora. For each instance, the sentence including the verb tense considered, as well as one sentence before or after, have been provided.

A way of evaluating human annotation is to calculate the inter-annotator agreement with the help of kappa coefficient (Carletta 1996). One issue that influences corpus annotation by raters is the subjectivity of the judgements, which can be quite important for semantic and pragmatic annotations (Artstein and Poesio 2008). 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. The kappa statistic factors out agreement by chance and measures the effective agreement by two or more raters. The kappa coefficient is 1 if
there is a total agreement among the annotators, 0 if there is no agreement other than the one expected to occur by chance, and -1 for values greater than chance agreement. We used this measure for quantifying the inter-annotator agreement in our experiments.

3.3.2 Annotation of conceptual information

Through this annotation experiment, we wanted to determine the conceptual meaning of two verb tenses in EN, SP and PresPerf. Our expectation was that human annotators should be able to think of the meaning of SP and PresPerf consciously, conceptualize it and make specific decisions in each context with facility. Annotators received annotation guidelines (presented below) and went through a training phase before the actual annotation phase. For the annotation of conceptual information, we expected high agreement related to the facility the task: easily graspable concepts and accessible to consciousness, while for procedural information we expected low agreement, related to a more difficult task: procedural information is rigid to conceptualization and not accessible to consciousness.

Based on our assumption (Grisot et al., submitted) that the configuration of Reichenbachian coordinates should be split in three pairs of two coordinates (E/R, R/S and the inferred E/S) instead of the classical view of three coordinates (as proposed by Reichenbach), we defined the conceptual content of the Simple Past as being the pair E<S which bears the focus (from the line E=R, R<S and E<S), in other words ‘situation that happened in the past’ and the conceptual meaning of Present Perfect as the pair R=S (from the line E<R, R=S, E<S), in other words the “current resulting state of a past situation”.

The annotation guidelines included: (a) a description of the two types of meaning (b) four examples for each usage, as given in the examples below and (c) the instruction to read each excerpt, identify the meaning of the verb highlighted and decide the type of meaning. In the first example, the most salient information is the result state in the present: the false declaration. In the second example, the most salient information is the situation that happened in the past: the lack of choice of Musharraf.

(1) And instead of full cooperation and transparency, Iraq has filed a false declaration to the United Nations that amounts to a 12,200-page lie.

(2) In a historic ruling that Musharraf had little choice but to accept, the Supreme Court itself reinstated the Chief Justice in July. Subsequently, the energized judiciary continued ruling against government decisions, embarrassing the government -- especially its intelligence agencies.

In what concerns the annotation guidelines, three aspects should be mentioned: (a) the ‘meaning’ of the SP and PresPerf, respectively, was easily identified and conceptualized in order to explain to annotators the task, (b) they were asked to identify ‘the most salient information’ in order to identify the focus and (c) annotators understood with easiness the annotation task, as well as the examples used for training.

In this experiment, annotators made decisions following the annotation instructions on 30 excerpts from the corpus. They agreed on the totality of the items annotated (kappa= 1) and pointed out the facility of the task. This result provides evidence for the conceptual nature of the information considered in this experiment.
3.3.3 Annotation of procedural information

One of the features tested with the help of the annotation experiment is \([\pm \text{narrativity}]\). As mentioned, this feature is a procedural information encoded by tenses that instructs the hearer/reader to verify if the reference point is part of a series of R that increases incrementally, or in other words if the eventualities presented are temporally ordered. If the feature is activated ([+ narrative]), then we can talk about a narrative usage of the verb tense considered. And respectively, if the feature is not activated, then the verb tense considered has a non-narrative usage.

Numerous studies have already addressed narrativity either in the traditional rhetoric (since the 19th century, such as Alexander Bain 1866 and John Genung 1900), in the DRT (Kamp and Reyle 1993) and SDRT (Lascarides and Asher 1993) or within a semantics and pragmatics perspective (Hinrichs 1986; Partee 1984; Reboul and Moeschler 1998a, 1998b; Smith 2001, 2003, 2010). Mainly, in these studies, narrativity is a discourse relation or a discourse mode associated with temporal sequencing of eventualities. In this chapter, we adopt this view of narrativity and postulate that it is a binary variable ([\pm \text{narrativity}]) that represents procedural pragmatic information conveyed by verb tenses and which can be used as a disambiguation criterion for various usages of tenses expressing past time in EN and FR.

The verb tense considered in this annotation experiment is the EN Simple Past. As in the first experiment, annotators received annotation guidelines (presented below) and went through a training phase. Narrativity has been defined and explained to annotators as it follows:

(3) In narrative contexts a story that is being told (you might not have the whole story available in the sentence) and eventualities are temporally ordered, while non-narrative contexts are associated with descriptive passages, where no story is being told.

Annotation guidelines included: (a) a definition of narrativity (b) the explanation of each usage (narrative and non-narrative) with two examples for each usage, as given in the examples below, (c) the instruction to read each excerpt, identify the verb highlighted and decide if in context, the highlighted verb is part of the underlying theme (the verb tense would have a narrative usage) or not (the verb tense would have a non-narrative usage).

In the first example 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.

(4) By his own marriage, likewise, which happened soon afterwards, he added to his wealth. (J. Austen, Sense and Sensibility)

(5) 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. (J. Austen, Sense and Sensibility)

The value of kappa coefficient for this annotation experiment was 0.42. This value is above chance, but not high enough to point to entirely reliable linguistic decisions (values generally accepted around 0.6- 0.7). What this first result shows about the procedural feature \([\pm \text{narrativity}]\) encoded by the EN SP is the difficulty hearers/readers have in the interpretation process to conceptualize the language rules they have and make decisions about their functioning.
The two annotators agreed on 325 items (71%) and disagreed on 133 items (29%). Error analysis showed that the main source of errors was the length of the temporal interval between two eventualities, which was perceived differently by the two annotators (that lead to ambiguity between temporal sequence or simultaneity, each of them corresponding to narrative, respectively, non-narrative usage). A possible explanation is the fact that personal world knowledge is used to infer temporal information, such as the length of the temporal interval between two eventualities, i.e. information that allows the annotator to decide whether the eventualities are temporally ordered or not. Cases where the length of the temporal interval between two eventualities was very reduced were ambiguous for the annotators, so each of them decided differently whether it was long enough for temporal sequencing or too short, so that the simultaneity meaning was preferred.

Disagreements were resolved in a second round of the annotation experiment, where the narrativity feature has been identified with a new linguistic test that was explained to two new annotators. Annotators were asked to insert a connective such as and and and then when possible, in order to explicit the ‘meaning’ of the excerpt, namely the temporal relation existent between the two eventualities considered. The connective because (for a causal relation) has also been proposed by annotators under the [+ narrative] label showing that causal relations should also considered. We will not look more into causality in this paper. The inter-annotator agreement in this second experiment was kappa = 0.91, signalling very strong and reliable agreement. This result emphasizes the procedural nature of the feature taking into account that one of the characteristics is the possibility to render explicit the instructions encoded with the help of discourse markers.

The cross-linguistic application of these findings consists of the observation of a pattern in parallel corpus. We investigated the data containing agreements of both annotators (325 items or 71% of the annotated data) and analyzed them in parallel corpus. We observed that the narrative usages of the SP identified by annotators correspond to narrative usages in the FR part of the corpus (translation by a PC or PS) in 152 instances, and the non-narrative usages of the SP correspond to the non-narrative usages in the FR text (translation with an IMP) in 111 instances (in total 80% within the data of agreements). This leaves 62 items where annotators agreed on the narrativity label but where it is not consistent with the tense used in FR. Future work will include an interchangeability test for these 62 items, where annotators agreed on narrativity labels but where the latter were not consistent with the tense usage in FR.

Because we worked only on a sample of items of the SP, we wanted to verify the generalization of our conclusions. This can be made with a statistical significance test, such as the chi2 test for independence. The result of the chi-square is chi2 = 119.50, df = 1, p < 0.001. The very low p-value allows us to reject the null-hypothesis and to

---

10 The new annotators were one of the authors and a research peer, who was not aware of the purpose of the research.

11 In Grisot et al. (submitted), we describe a similar annotation experiment made on the French tenses used for translating the EN SP, namely PC, PS and IMP. In this experiment, the PC and PS have been identified as being narrative and the IMP as being non-narrative with a kappa value of 0.63 (reliable agreement).

12 Gries (2013) explains that for two categorical variables, the chi-square test for independence helps to determine whether the values of the independent variable are correlated with those of the dependent variable.
conclude that the observed correlation between the type of usage of SP and the tense used for its translation in FR is extremely unlikely to occur by chance. In order to quantify this correlation effect, we have further computed a correlation coefficient called Cramer’s V (cf. Gries 2013) that ranges from 0 to 1, where the larger the value, the stronger the correlation. In our data, we obtained a value of 0.61, which points to strong correlation.

3.4 NLP application

Nowadays, linguistic research tends more and more to integrate language automatic processing techniques. Human annotation and classification of texts is often used in Natural Language Processing (NLP) domain. Human-annotated texts serve as examples or as training items for machine-learning tools. The annotated data from our case study was used\textsuperscript{13} as data for automatic annotation with a Maximum Entropy (MaxEnt) classifier with the Stanford Classifier Package (Manning and Klein 2003). A classifier is a machine-learning tool that will take data items and place them into one of the available classes. The underlying principle of maximum entropy is that, when assigning a class if there is no external knowledge, one should prefer uniform distributions, thus assign uniformly the considered classes. Annotated data used for training these classifiers provide external knowledge and thus inform the automatic labelling technique where to be minimally non-uniform. Iterative runs of the classifier results in automatically labelled or annotated texts with the considered features.

The feature tested in our case study was \([\pm \text{narrativity}]\) and the human-annotated data was used for training the classifier. The results of automatic annotation are similar to human annotation; the classifier correctly annotated 76% of the items. The purpose of using automatic annotation is the possibility to do it on large amounts of data. Human annotation has the disadvantages of being tedious and costly, and it is often done on a reduced amount of data.

The final purpose was to improve the results in what concerns verb tenses of a statistical machine translation system. Current statistical machine translation (SMT) systems have difficulties in choosing the correct verb tense translations, in some language pairs, because these depend on a wider-range context than SMT systems consider. Indeed, decoding for SMT is still at the phrase or sentence level only, thus missing information from previously translated sentences (which is also detrimental to lexical cohesion and co-reference).

A first run of the SMT system used based on the annotated data with the \([\pm \text{narrativity}]\) feature had slightly better results than without this pragmatic feature. When trained and tested on the automatically annotated data, the \([\pm \text{narrativity}]\) feature improves translation by about 0.2 BLEU points. More importantly, manual evaluation shows that verb tense translation and verb choice are improved by respectively 9.7% and 3.4% (absolute), leading to an overall improvement of verb translation of 17% (relative) (for more detailed results see Meyer et al. 2013).

4 Conclusion

This paper has given an account of the place of empirical pragmatics among theoretical pragmatics and experimental pragmatics, for the study of language in use.

\textsuperscript{13} The NLP work was done by our colleagues Thomas Meyer and Andrei-Popescu Belis from the Idiap Research Institute (Martigny, Switzerland) to whom we address our gratitude.
We have argued for the need to have robust data for pragmatic research, data provided by both corpus work and experimentation.

We have shown that corpus work can be fruitfully done with a contrastive perspective, following the specific three-steps methodology of Contrastive Analysis domain. In what concerns experimentation, we have looked into offline experiments consisting of linguistic judgement task, experiments that in the NLP domain are considered as human annotation experiments. We have discussed the example of the first experiment for the pragmatic distinction between what is ‘said’ and what is ‘implicated’ designed by Gibbs and Moise (1996). Another important topic of this paper was the discussion about the advantages and difficulties of each of the two methods considered (corpus work and experimentation), as well as their complementarity.

In our case study, we investigated the nature of the information encoded by verb tenses. We assumed and validated empirically through annotation experiments that verb tenses encode both procedural and conceptual information. We defined conceptual information as being involved in the language of thought in a Fodorian framework (Fodor 1975, 1998) having the characteristic of being accessible to consciousness and capable of being reflected on, evaluated and used in general inference. We proposed thus, based on these two features, that verb tenses encode conceptual information consisting of a certain configuration of temporal coordinates. The basic meaning of a tense is to locate an eventuality related to the speech moment, passing through a reference point.

In what concerns procedural information, we followed Wilson and Sperber’s idea (1993) that procedures are not part of language of thought and thus are not accessible to consciousness and easily conceptualized, as representations are. The results of the annotation experiment showed that verb tenses encode procedural information that instruct the reader/hearer to look for other eventualities that are related to the eventuality considered, namely the [± narrativity] procedural feature.

Taken together, the empirical findings of this research provide an example of the relation between theoretical framework(s) and empirical methodologies. Theoretical hypotheses have an impact on the choice of empirical methodologies, for example, the cross-linguistic perspective when investigating the translation of verb tenses requires work on parallel corpora in order to have access to both source and target texts. The disambiguation of the usages of the targeted verb tense requires the formulation of possible disambiguation criteria that need to be validated through experimentation involving linguistic judgement tasks. In what concerns the impact of genuine data dealt with empirical methods (qualitative and quantitative) on theoretical hypotheses, we have shown that results from experimental work can challenge some theoretical positions. For verb tenses, for example, the results of our experiments challenged the theoretical assumption that verb tenses do not encode conceptual information, but only procedural information.

Finally, our work has illustrated how empirical pragmatics pools resources with the NLP domain. The pragmatic feature identified as procedural information and validated through human annotation experiments, has been used as label for discourse tagging with an automatic classifier. Moreover, an SMT system trained on the annotated corpus had better results for translating verb tenses than if it hadn’t made use of the [± narrativity] pragmatic feature.

An issue that was not addressed in this study was the cross-linguistic application of the model to more than one pair of languages. This issue will be addressed in further studies and it targets the translation of the English Simple Past in Italian and
Romanian, next to French. The current study was not specifically designed to investigate the conceptual/procedural distinction and its application for verb tenses using online experimental methodology. This would be very interesting and probably, it will fade any doubts about existent conceptual content of verb tenses.

Bibliography


Bain, Alexander. 1866. An English Grammar. Longmans, Green, Reader, and Dyer.


Smith, Carlota S. 1990. « From graduate school to tenure ». The Cornell Lectures: Women in the linguistics profession: 239–244.


