Towards a lexicon of event-selecting predicates for a French FactBank

Ingrid Falk and Fabienne Martin
University of Stuttgart
`firstname.lastname@ling.uni-stuttgart.de`

Abstract

This paper presents ongoing work for the construction of a French FactBank and a lexicon of French event-selecting predicates (ESPs), by applying the factuality detection algorithm introduced in (Saurí and Pustejovsky, 2012). This algorithm relies on a lexicon of ESPs, specifying how these predicates influence the factuality of their embedded events. For this pilot study, we focused on French factive and implicative verbs, and capitalised on a lexical resource for the English counterparts of these verbs provided by the CSLI Group (Nairn et al., 2006; Karttunen, 2012).

1 Introduction

Texts not only describe events, but also encode information conveying whether the events described correspond to real situations in the world, or to uncertain, (im)probable or (im)possible situations. This level of information concerns *event factuality*. This study reports ongoing work on the annotation of French TimeBank events with event factuality information, whose main goal is the elaboration of a French FactBank. We plan to achieve this as follows. Saurí and Pustejovsky (2012) developed an elaborate model of event factuality that allows for its automatic detection. We aim to capitalise on this work by applying Saurí and Pustejovsky (2012)'s algorithm to the events in the French TimeBank (FTiB henceforth) and assign to these a factuality profile. Given that the FTiB has only about 1/4 of the size of the English TimeBank, we will manually review the automatically obtained factuality profiles in a second step.

Saurí and Pustejovsky (2012)'s algorithm relies on two crucial prerequisites. The first is the identification of the *sources* at play, i.e. the cognitive agents endorsing a specific epistemic stance on the events described. The text author is the default source, but some linguistic constructions — and a subclass of verbs in particular, see e.g. *affirm* — present one or more sources that are also committed to the factuality of the reported event. Secondly, the algorithm makes use of three language-specific and manually developed *lexical resources*, capturing the way polarity particles, particles of epistemic modality and so-called event-selecting predicates (e.g. *manage to, suspect that*) influence event factuality. In this study, we show how existing lexical semantic resources can be used and modified in order to build the French-specific lexical resources needed to apply Saurí and Pustejosvki's algorithm to the French TimeBank.

2 The English FactBank

As described in (Saurí and Pustejovsky, 2009), FactBank is an English corpus annotated with information concerning the factuality of events. It is built on top of the English TimeBank by adding a level of semantic information. TimeBank is a corpus annotated with TimeML (Pustejovsky et al., 2005), a specification language representing temporal and event information in discourse. The factuality information encoded in TimeBank and relevant for our work are the event-selecting predicates (cf. Section 3) which project a factual value to the embedded event by means of subordination links (or SLINKs).

In TimeBank, a total of 9 488 events across 208 newspaper texts have been manually identified and annotated. FactBank assigns additional factuality information to these events. More specifically, it is annotated for each event (i) whether its factuality is assessed by a source different from the text author and (ii) the degree of factuality the new source and the text author attribute to the event (for in-

stance, *Peter affirmed P* presents *P* as certain for Peter, but does not commit the text author to *P* in a specific way). Saurí and Pustejovsky (2009) distinguish six 'committed' factuality values (i.e. values to which a source is committed) and one 'uncommitted' value, which are shown in Table 1.

Saurí and Pustejovsky (2012) present an algorithm which assigns to each TimeBank event a factuality profile consisting of (i) its factuality value, (ii) the source(s) assigning the factuality value to that event and (iii) the time at which the factuality value assignment takes place. The algorithm assumes that events and relevant sources are already identified and computes the factuality profile of events by modelling the effect of factuality relations across levels of syntactic embedding. It crucially relies on three lexical resources which the authors developed manually for English. Since to apply this algorithm to the French data we need to create similar resources for French, we describe them in more detail in the following section.

3 Lexical Resources for the Automatic Detection of Factuality Profiles

The first lexical resource is a list of 11 negation particles (adverbs, determiners and pronouns) which determine the polarity of the context, together with a language independent table showing how these polarity markers influence the polarity of the event. The corresponding list of negation particles needed for French can be set up easily.

The second resource aims to capture the influence of epistemic *modality* on the event. It gives a list of 31 adjectives, adverbs and verbs of epistemic modality together with the factuality value they express. Most of their French counterparts influence the context the same way as in English, except for modal verbs, that are well-known to give rise to an 'actuality entailment' under their root/non-epistemic readings (i.e. to present the embedded event as a fact in the real world) when combined with a perfective tense, see (Hacquard, 2009), an issue briefly addressed in Section 4.

The third resource is the most complex one and accounts for the influence on the event factuality value in cases where the event is embedded by so-called *event-selecting predicates (ESPs)*, of which *suspect that/manage to* are examples. ESPs are predicates with an event-denoting argument, which lexically specify the factuality of the event. Saurí and Pustejovsky distinguish two kinds of

ESPs: *Source Introducing Predicates (SIPs)* introduce a new source in discourse (e.g. *suspect/believe*); *Non Source Introducing Predicates (NSIPs)* do not (e.g. *manage/fail*). As part of their lexical semantics, SIPs determine (i) the factuality value the new source (the 'cogniser') assigns to the event described by the complement, and (ii) the factuality value assigned by the text author (i.e. the 'anchor') to the same event. NSIPs, on the other hand, determine event factuality wrt. a unique source, the anchor. In addition, the assessment of event factuality wrt. the relevant source(s) varies with the polarity and modality present in the context of the ESP. Table 1 illustrates the lexicon layout through sample entries for the NSIPs *manage* and *fail*.[1] The ESP lexicon built by Saurí and Pustejovsky (2012) consists of 646 entries in total (393 verbs, 165 nouns and 88 adjectives). In order to apply Saurí and Pustejovsky (2012)'s algorithm to the French TimeBank, we need to build a similar ESP lexicon for French. To speed up this process, we plan to use a large body of research about English predicates with sentential complements (Karttunen, 1971b; Karttunen, 1971a; Nairn et al., 2006; Karttunen, 2012)[2], which we briefly introduce in the following.

Factive and implicative verbs. Nairn et al. (2006) develop a semantic classification of complement-taking verbs according to their effect on the polarity of their complement clauses. This classification is shown in Table 2. We illustrate how the table works in the following examples. In example (1), the ESP *fail to* has positive polarity. We obtain the factuality of the embedded event (*reschedule*) by retrieving from the polarity $+$ column in Table 2 the polarity value in the *fail to* row, which is '$-$', i.e. the meeting is not rescheduled (has factuality CT$-$). For (2), the factuality must be retrieved from the polarity $-$ column resulting in '$+$', i.e. a factuality of CT$+$ (the meeting is rescheduled).

(1) Kim *failed to* reschedule the meeting.

(2) Kim did not *fail to* reschedule the meeting.

The effect of a predicate on the polarity of its embedded complement is represented more concisely

---

[1] The lexicon layout for SIPs, less relevant for our study, is very similar except that an SIP lexicon entry must also provide factuality values for the cogniser source in addition to the anchor.

[2] The ESP lexicons created in (Saurí and Pustejovsky, 2012) are not freely available.

Contextual factuality												
	CT			PR			PS			U		
polarity	+	−	u	+	−	u	+	−	u	+	−	u
manage (a)	CT+	CT−	CTu	PR+	PR−	PRu	PS+	PS−	PSu	Uu	Uu	Uu
fail (a)	CT−	CT+	CTu	PR−	PR+	PRu	PS−	PS+	PSu	Uu	Uu	Uu

Table 1: Sample lexical entries for NSIPs. CT, PR and PS signify certain, probable and possible respectively, U (and/or u) unspecified (unknown or uncommitted), (a) refers to the anchor.

	Polarity of ESP		Sample
	+	−	predicate
2-way	+	−	*manage to*
implicatives	−	+	*fail to*
1-way	+	n	*force to*
+implicatives	−	n	*refuse to*
1-way	n	−	*attempt to*
-implicatives	n	+	*hesitate to*
factives	+	+	*forget that*
counterfactives	−	−	*pretend that*
Neutral	n	n	*want to*

Table 2: Semantic classification of complement taking verbs wrt. the effect of the polarity of the main clause (ESP, head row) on the factuality of the complement clause (embedded event, subsequent rows). n stands for none.

through a "signature". For instance, the signature of factive verbs as *forget that* is '$++|-+$' (Read: 'if positive polarity, event happens; if negative polarity, event happens').

Thus, based on the signature of a predicate and its polarity in a given sentence, we can determine the factuality of the embedded event in that sentence. It should now be obvious how the classification in Table 2 can be "plugged" into the ESP lexical resources illustrated in Table 1:[3] For a given ESP for which a lexical entry has to be set up (*eg. fail*), the factuality value conveyed on the embedded event can be retrieved from Table 2 whenever the corresponding table entry is not n. In case it is, the polarity value must be set to u (unspecified).

Nairn et al. (2006) compiled a list of roughly 250 English verbs found to carry some kind of implication: a positive or negative entailment, a factive or a counterfactive presupposition[4]. To test how this approach can help us build the French ESP lexical resource required for a French factuality profiler, we translated these English verbs into

---

[3]Factive and implicative verbs are typically non-source introducing predicates (NSIPs).

[4]These resources are available at `https://web.stanford.edu/group/csli_lnr/Lexical_Resources/`.

their French counterparts, and looked at the sentences in the French TimeBank using these French counterparts as ESPs. We first briefly introduce the French TimeBank before describing our data, experiments and findings.

4 Experiments on the French TimeBank

The French TimeBank (Bittar, 2010; Bittar et al., 2011) is built on the same principles as the English TimeBank, but introduces additional markup language to deal with linguistic phenomena not yet covered and specific to French. Most relevant to this study are the following. FTiB uses ALINK elements to encode aspectual subordination relations, holding between events realised by an aspectual verb (e.g. *commencer* 'begin') and a subordinated event complement. The subordinating events in ALINKs, as those in SLINKs, are ESPs and are therefore also relevant for this study. Also, since French modal auxiliaries can be fully inflected and fall within the scope of aspectual operators, they are also marked up as events. Lastly, the TimeML schema was adapted to represent the grammatical tense/aspect system of French, and to account *eg.* for the *imparfait* (IMPERFECT), not grammaticalised in English.

FTiB is made up of 108 newspaper texts for a total of 16 208 tokens. 2 098 of these represent events. Since in our experiments, we are interested in assessing factuality at the sentence level, we segmented FTiB into (814) sentences, and extracted from them the subordination links (SLINKs and ALINKs). Overall, FTiB contains 485 subordination links. Luckily, the subordinating events in 444 of these links are ESPs. From these links, we selected those where the subordinating event was a translation of an English verb for which we have a signature (179, instantiating 49 different verbs). We first checked for the 49 types whether the French predicate has the same signature as the English verb it translates. We found that this was very roughly the case for most of

these verbs (but see below). For example, the factive signature '$+ + | - +$' of *learn that* is inherited by its translation *apprendre que*. Similarly, the implicative signature '$+ - | - n$' of *help* also characterises its French translation *aider à*. Our translation approach raised several interesting issues, however. A first one concerns verbs with a syntax-dependent factuality profile. In English, *learn that/forget that* for instance have the (factive) signature $+ + | - +$, whereas *learn to* is less biased wrt. the factuality of the embedded event, and *forget to* has the (implicative) signature $+ - | - +$. French translations of these verbs also quite systematically see their signatures varying with the syntactic structure, or have an argument structure that its English counterpart cannot instantiate; see e.g. the factive VP *apprendre sa mort* (lit. '*learn his death'). For these cases, we paired the relevant reading with the appropriate signature manually. A second issue is raised by verbs with an aspect-dependent factuality profile. It is well-known that in Romance, modal verbs trigger an actuality entailment under some of their readings, but *with a perfective only*. For instance, the example (3), where the modal verb *permettre* has an *enable* reading and is combined with a perfective (PFV), triggers an actuality entailment (the embedded event has to happen). With an imperfective (IMP), however, the actuality entailment vanishes, cf. (4). Also, when the same verb *permettre* has a deontic ('grant permission') reading, no actuality entailment is triggered, even with a perfective, see (5) (Hacquard 2006:41).

(3) *Cette carte **m'a permis** d'entrer.*
    This card me permit-PFV to enter
    → I entered.

(4) *Cette carte **me permettait** d'entrer.*
    This card me permit-IMP to enter
    ↛ I entered.

(5) *Le doyen **m'a permis** d'entrer.*
    The dean me permit-PFV to enter
    ↛ I entered.

Our translations of the English verbs analysed by Nairn et al. (2006) revealed that several other French verbs show the same lability, see Table 3. That is, the entailment triggered with the perfective is lost with an imperfective, or at least replaced by a defeasible inference, see e.g. the examples (6)-(7).

	ASPECT USED			
	PFV		IMP	
	Polarity of ESP			
	+	−	+	−
assurer (la victoire) insure (the victory)	+	n	n	n
condamner (x à rester) condemn (x to stay)	+	n	n	n
conduire (à la catastrophe) lead to (catastrophy)	+	n	n	n
apprendre (à voler) learn (to fly)	+*	−*	n	n
réussir (à entrer) manage (to enter)	+	−	n	n
daigner (répondre) deign (to answer)	+	−	n	n
motiver (x à venir) motivate (x to come)	+*	−*	n	n
échouer (à persuader x) fail (to persuade x)	−	+	n	n

Table 3: Examples of verbs whose inferential profile varies with the aspect used. Certain events are labelled '+', very likely but not certain events, '+*', counterfactual events, −, very unlikely events '−*'.

(6) *A ce moment-là, elle **a réussi** à*
    at that moment she manage-**PFV** to
    s'enfuir. #Mais finalement, elle
    escape but finally she
    ne s'est pas enfuie.
    NEG escape-PFV
    'At that moment, she managed to escape. But finally, she didn't escape.'

(7) *A ce moment-là, elle **réussissait***
    at that moment she manage-**IMP**
    (encore) à s'enfuir. OK Mais
    (still) to escape but
    finalement, elle ne s'est pas enfuie.
    finally she NEG escape-PFV
    'At that moment, she 'was managing' to escape. But finally, she didn't escape.'

Interestingly, most of these predicates with an aspect-dependent inferential profile (12 out of 13 in the current stage of annotation) are implicative verbs. On the other hand, verbs whose inferential profile is independent from aspect are mostly factive ($+ + | - +$) verbs (17 out of 23). The verb *savoir* illustrates well the point. Used as a translation of the English factive verb *know*, *savoir* is factive both with PFV and IMP. However, *savoir* is also used in the FTiB as a two-way implicative verb ($+ + | - -$), see (8)-(9). In the latter use,

savoir has an abilitative reading and like *être capable de* 'be able to', triggers an actuality entailment with PFV, see (8), but has a neutral inferential profile with IMP ($+n| - n$), see (9).

(8) *Il **a su** maîtriser ce fantasme.*
he be able-**PFV** master this fantasy
→ He mastered this fantasy.

(9) *Il **savait** maîtriser ce fantasme.*
he be able-**IMP** master this fantasy
↛ He mastered this fantasy.

Why do implicative verbs (contrary to factive verbs) lose their entailment when combined with IMP? Recent analyses of implicative verbs by (Baglini and Francez, 2016) and (Nadathur, 2016) can help to explain this observation. According to Baglini & Francez' analysis, a *manage p* statement *presupposes* familiarity with a causally necessary but insufficient condition *A* for the truth of *p*, and *asserts* that *A* actually caused the truth of *p*. Nadathur extends a modified version of this analysis to the whole class of implicative verbs. The important point for us is that under these analyses, implicative verbs have an at-issue component: they assert an 'event', namely the obtaining/actualisation of the causal factor *A* for the truth of *p*. Given the 'imperfective paradox', the imperfective form of such verbs unsurprisingly *suspends* the actualisation event, as what happens with the imperfective form of overtly causative verbs (e.g. *Trump was causing a new catastrophe when Pence stopped him* does not entail the occurrence of a new catastrophe). On the other hand, factive verbs like *savoir que p* 'know that *p*' do not assert the obtaining of a causal factor for the truth of *p*, but rather a mental state having *p* as its object. We therefore do not expect aspect to interfere with their inferential profile.

For these verbs with an aspect-dependent aspectual profile (including plainly modal ones)[5], we manually annotated the reading instantiated and the corresponding signature in the FTiB.

The third interesting point raised by our translation is illustrated by French verbs having a different factuality profile than their English counterparts. For instance, although *pousser à* is used to translate the implicative verb *provoke to*, it is not implicative with an agent subject, even with a perfective (contrary to its near synonym *conduire à*).

---

[5]FTiB has only 32 instances of *devoir*, 8 of *falloir* and 21 of *pouvoir*.

5 Ongoing research

These experiments showed that verbs whose factuality profile varies with the reading selected and/or its argument structure are very pervasive among French ESPs. A lexicon of ESPs should therefore carefully distinguish between the different readings/argument structures an ESP may instantiate. Also, they suggest that interesting new correlations can be found between event factuality profiles on one hand, and particular sets of syntactic/semantic properties on the other. For instance, verbs like *refuser* 'refuse/fail' are two way implicative verbs with an inanimate subject or with an animate subject controlling the complement, cf. (10)-(11), but only trigger a strong (but nevertheless defeasible) inference with a matrix subject distinct from the infinitival subject, see (12).

(10) *Le tiroir **a refusé de** s'ouvrir, #mais il s'est ouvert quand même.*
'The drawer failed to open, but it opened nevertheless.'

(11) *Marie **a refusé d'**entrer, #mais elle est entrée quand même.*
'Marie refused to enter, but she entered nevertheless.'

(12) *Le garde **a refusé que** Marie entre, OK mais elle est entrée quand même.*
'The guard refused to allow Marie to enter, but she entered nevertheless.'

To find these correlations, we are building a French lexicon of ESPs on top of a rich lexicon encoding morphological, syntactic and semantic properties of French verbs *for each of their readings*, "Les verbes français" (Dubois and Dubois-Charlier, 1997; François et al., 2007). In the first step, we use the French verbs analysed in our experiments as seeds, link them with each of their readings in *Les verbes français*, and provide a manual signature for all of their other ESP readings. This will hopefully give an idea of the semantic and syntactic properties characterising each factuality profile. In the second step, we will enrich the different subclasses of ESPs (distinguished by their signature) with similar candidates by using (semi-)automatic methods along the lines of those described in (Richardson and Kuhn, 2012; De Melo and De Paiva, 2014; White and Rawlins, 2016; Eckle-Kohler, 2016), and then review them manually.

References

Rebekah Baglini and Itamar Francez. 2016. The Implications of Managing. *Journal of Semantics*, 33/3:541–560.

André Bittar, Pascal Amsili, Pascal Denis, and Laurence Danlos. 2011. French TimeBank: an ISO-TimeML annotated reference corpus. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 130–134. Association for Computational Linguistics.

André Bittar. 2010. *Building a TimeBank for French: a reference corpus annotated according to the ISO-TimeML standard*. Ph.D. thesis, Paris 7.

Gerard De Melo and Valeria De Paiva. 2014. Sense-specific implicative commitments. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 391–402. Springer.

Jean Dubois and Françoise Dubois-Charlier. 1997. *Les Verbes français*. Larousse.

Judith Eckle-Kohler. 2016. Verbs Taking Clausal and Non-Finite Arguments as Signals of Modality - Revisiting the Issue of Meaning Grounded in Syntax. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*.

Jacques François, Denis Le Pesant, and Danielle Leeman. 2007. Présentation de la classification des Verbes Français de Jean Dubois et Françoise Dubois-Charlier. *Langue française*, 153(1):3–19.

Valentine Hacquard. 2009. On the Interaction of Aspect and Modal Auxiliaries. *Linguistics and Philosophy*, 32:279–312.

Lauri Karttunen. 1971a. Implicative Verbs. *Language*, 47(2):340–358.

Lauri Karttunen. 1971b. *The logic of English predicate complement constructions*, volume 4. Indiana University Linguistics Club Bloomington, Indiana.

Lauri Karttunen. 2012. Simple and phrasal implicatives. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 124–131. Association for Computational Linguistics.

Prerna Nadathur. 2016. Causal necessity and sufficiency in implicativity. In *Proceedings of Semantics and Linguistics Theory (SALT) 26*, pages 1002–1021.

Rowan Nairn, Cleo Condoravdi, and Lauri Karttunen. 2006. Computing relative polarity for textual inference. In *Proceedings of the Fifth International workshop on Inference in Computational Semantics (ICoS-5)*, pages 20–21.

James Pustejovsky, Robert Knippen, Jessica Littman, and Roser Saurí. 2005. Temporal and event information in natural language text. *Language resources and evaluation*, 39(2):123–164.

Kyle Richardson and Jonas Kuhn. 2012. Light textual inference for semantic parsing. In *24th International Conference on Computational Linguistics*, page 1007.

Roser Saurí and James Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227.

Roser Saurí and James Pustejovsky. 2012. Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299.

Aaron Steven White and Kyle Rawlins. 2016. A computational model of s-selection. In *Semantics and Linguistic Theory*, volume 26, pages 641–663.