# Investigating ‘Aspect’ in NMT and SMT

**Article** *in* Computational Linguistics in the Netherlands Journal · November 2017



1 author:

Eva Vanmassenhove
Dublin City University

# Investigating 'Aspect' in NMT and SMT:
## Translating the English Simple Past and Present Perfect

**Eva Vanmassenhove** EVA.VANMASSENHOVE@ADAPTENTRE.IE
**Jinhua Du** JINHU.DU@ADAPTCENTRE.IE
**Andy Way** ANDY.WAY@ADAPTCENTRE.IE

*Adapt Centre, Dublin City University, Dublin, Ireland*

## Abstract

One of the important differences between English and French grammar is related to how their verbal systems handle aspectual information. While the English simple past tense is aspectually neutral, the French and Spanish past tenses are linked with a particular imperfective/perfective aspect. This study examines what Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) learn about 'aspect' and how this is reflected in the translations they produce. We use their main knowledge sources, phrase-tables (SMT) and encoding vectors (NMT), to examine what kind of aspectual information they encode. Furthermore, we examine whether this encoded 'knowledge' is actually transferred during decoding and thus reflected in the actual translations. Our study is based on the translations of the English *simple past* and *present perfect* tenses into French and Spanish imperfective and perfective past tenses. We examine the interaction between the lexical aspect of English simple past verbs and the grammatical aspect expressed by the tense in the French/Spanish translations. It results that SMT phrase-tables contain information about the basic lexical aspect of verbs. Although lexical aspect is often closely related to the grammatical aspect expressed by the French and Spanish tenses, for some verbs (mainly atelic dynamic verbs) more contextual information is required in order to select an appropriate tense. The SMT $n$-grams provide insufficient context to grasp other aspectual factors included in the sentence to consistently select the tense with the appropriate aspectual value. On the other hand, the encoding vectors produced by our NMT system do contain information about the entire sentence. An analysis based on the English NMT encoding vectors shows that a logistic regression model can obtain an accuracy of 90% when trying to predict the correct tense based on the encoding vectors. However, these positive results are not entirely reflected in the actual translations, i.e. part of the aspectual information is lost during decoding.

## 1. Introduction

Translating sentences from one language to another is a complex task that requires a profound knowledge of the two languages involved. Translators use their understanding of the morphology, structure and the semantics of both languages in order to select an appropriate translation in a specific context (Hogeweg et al. 2009). Additional translation difficulties arise when dealing with "translation mismatches", a term used in the field of MT to refer to cases where the grammar of one language makes distinctions that are not made by the other. Such a mismatch can apply to a particular utterance or it can be due to more systematic differences between the source and target language systems.

Systematic differences between two languages reveal something about the linguistic systems involved. A better understanding of the systematicity of apparent mismatches between languages and the mechanisms behind them could lead to a more accurate mapping between two language systems. Although "two languages are more informative than one" (Dagan et al. 1991), not many corpus-based translation studies focus on specific linguistic phenomena (Santos 2004). In the field of MT, and more particularly Rule-Based MT (RBMT), comprehensive and detailed modules were integrated in MT systems (such as Eurotra and Rosetta) that dealt with mismatches related to tense and aspect (Van Eynde 1988, Appelo 1993). However, in data-driven (SMT and NMT), not many studies focus on resolving particular translation problems.

The mismatches that occur within the verbal systems of languages are particularly interesting since verbs are arguably the most important lexical category of language (Miller and Fellbaum 1991). Sentences are governed by verbs and, with the exception of some languages such as Russian (where it is possible to have a sentence that does not contain a verb), languages need verbs to represent the sentence predicate (Čech et al. 2011). Furthermore, verbs have a crucial impact on the general structure of sentences and are the most complex and varied forms in language (Fischer and Gough 1978). Thus, incorrect translations of verbs can propagate errors across a sentence.

English and French/Spanish grammar have considerable areas of overlap, but there are some important mismatches that can cause interference. French and Spanish have, for example, a richer inflectional morphology where verbs need to agree in person and number with their subject (or objects in some cases). Although this is a relatively easy task for human translators, these translation mismatches appear to be difficult to learn for SMT systems (Vanmassenhove et al. 2016). There is also no one-to-one correspondence between the English and French/Spanish tense systems. Some tenses are formally similar, such as the English *present perfect* and the French *passé composé*, but are not in usage.[1] In the case of the *present perfect* and the *passé composé*, this is due to a shift that occurred in the French past tenses system, where the *passé composé* has largely replaced the usage of the French *passé simple* and is thus used as a perfective past tense. Although a similar evolution has been observed in English, where the *present perfect* is used increasingly in *simple past* contexts (presumably by analogy with the French tense), the *present perfect* forms cannot express past events to the same extent as the *passé composé* (Engel 1998).

One of the main differences between the English and French/Spanish tenses is related to 'aspect' and 'tense'. Comrie (1976, p.3) defines tense and aspect as follows: "[...] tense relates the time of the situation referred to some other time, whereas aspects are different ways of viewing the internal temporal constituency of a situation." The English *simple past* is aspectually vague when compared to the French (*passé composé* and *imparfait*) and Spanish (*pretérito indefinido* and *pretérito imperfecto*) past tenses. Although both English and French/Spanish tenses express mainly the tense (present, past or future), the *passé composé/pretérito indefinido* and the *imparfait/pretérito imperfecto*, express the same tense (past) but have a different aspectual meaning. Their distinction is purely based on an aspectual difference and is thus an example of a formally marked aspect that forms part of the morphological system of French/Spanish, a "grammatical aspect" (Garey 1957). An example of the grammatical aspect expressed by the Spanish past tenses is given in Example 1.

**Example 1**

| | | | |
|---|---|---|---|
| *(a)* | *"He made dinner."* | "Hizo *la comida*" | *PERFECTIVE* |
| *(b)* | *"He made dinner."* | "Hacía *la comida*" | *IMPERFECTIVE* |

The perfective aspect expressed in Spanish by the *pretérito indefinido* is bounded. It presents the event from the outside, as a whole (Vendler 1957, Comrie 1976, Dowty 1986). The perfective reading of (a) "He made dinner" is one where the event started and finished (it has been 'completed') and thus resulted in a 'dinner'. However, (b) presents the event from the inside (Comrie 1976, Dowty 1986, Vendler 1957), without emphasis on the beginning or the end, merely focusing on the fact that 'He was making dinner'.

This does not imply that the English language does not or cannot convey this aspectual difference since other words in a sentence (the semantics of the verbs, nouns, adjectives, adverbs, etc.) can carry aspectual meaning. While some languages have an overt formal category of grammatical aspect, for others it is a covert semantic category on the sentential (or propositional) level (Filip 2012). The difficulty consists thus of making something that is implicit in the one language explicit when

1. The difference can already be noted by looking at their denominations: the 'present' perfect vs the 'passé' composé ('past' composed) tense.

translating into another language. This is a recurring problem in SMT occurring across languages: e.g. pronoun-dropping in Chinese-English (Wang et al. 2016).

As different words in a sentence can carry aspectual meaning, the aspect of a sentence should be regarded as a network. Within this network, the basic 'lexical aspect' of the verb is a good starting point to determine the aspectual value of a sentence/proposition (Moens 1987). Within the lexical structure of a verb, there can be properties that present some boundary or limit (with respect to the duration of the event described by the verb) while this is not the case for others verbs. A verb like 'to explode' or 'to sneeze', would, without any further context, be translated with a perfective tense since the action is completed the moment it happens. Other verbs like 'to own', 'to want' do not put such an emphasis on the beginning or the end of an event and are more easily linked with an imperfective aspect. In Example 2 we provide some example sentences in French that illustrate this interaction between lexical aspect and grammatical aspect.

**Example 2**

| | | | |
|---|---|---|---|
| *(a)* | *"It exploded."* | *"Ça* ***a explosé****"* | *PERF.* |
| *(b)* | *"It exploded* **continuously***".* | *"Ça* ***explosait*** *sans cesse"* | *IMPERF.* |
| *(c)* | *"I owned a car."* | *"Je* ***possédais*** *une voiture"* | *IMPERF.* |
| *(d)* | *"I owned a car* ***for two years****."* | *"J'****ai possédé*** *une voiture pendant deux ans"* | *PERF.* |

Whereas the English verb 'exploded' in (a) is prototypically linked with a perfective interpretation (and thus translated in French with a *passé composé*), the verb 'owned' in (c) has semantic properties that are more closely related to the imperfective aspect. However, an adverb (*continuously* in (b)) or prepositional phrase (*for two years* in (d)) can change the overall aspectual value of the proposition. The example sentences in (2) illustrate that 'grammatical aspect' and 'lexical aspect' are intertwined.

Verbs can be grouped together in a taxonomy of aspectual classes according to properties related to their 'lexical aspect'. Wilmet's taxonomy (Wilmet 1997) classifies verbs into three different aspectual classes: *stative* verbs (e.g. 'to own', 'to love', 'to believe', 'to want'), *telic*[2] *dynamic* verbs ('to cough','to deliver') and *atelic dynamic* verbs (e.g. 'to eat', 'to write', 'to walk'). Unless the context contains important triggers that suggest otherwise, stative verbs will be more likely translated into an imperfective tense. Likewise, telic activity verbs are prototypically translated into a perfective tense. However, this classification is not fixed. It is susceptible to aspectual triggers provided by the context. That is, verbs can transition from one class to another given the right triggers in the context (as illustrated in Example 2). This implies that an automatic translation system might need the entire source sentence (or even its surrounding sentences) in order to determine correctly the aspectual value of a verb.

Phrase-Based Statistical Machine Translation (PBSMT)[3] (Koehn et al. 2007) learns to translate phrases ($n$-grams) of the source language to target language phrases based on their co-occurrence frequencies in a parallel corpus. The size of those $n$-grams is usually limited to 6. All source language phrases and their target language counterparts are stored in phrase-tables together with their probabilities. Every phrase is seen as an atomic unit and thus translated as such. Given that these units are limited in length and not linguistically motivated but purely based on statistics, important linguistic information can be lost. When it comes to determining the aspectual value of a verb/proposition in English, SMT can, on the source side, rely only on the limited information contained in those phrase-tables.

Neural Machine Translation (NMT) (Bahdanau et al. 2015, Cho et al. 2014, Sutskever et al. 2014) encodes the entire source sentence in an encoding vector. In an encoder-decoder NMT model, there are two neural networks at work. The first one encodes information about the source sentence into a vector of real-valued numbers (the hidden state). The second neural network decodes the

2. Telic verbs are verbs that refer to events/situations with an inherent goal or ending while atelic verbs do not have this property (Garey 1957).
3. For simplicity, we will refer to PBSMT as SMT in the remainder of this paper.

hidden state into a target sentence. Unlike SMT, the neural network responsible for the decoding of the hidden state has access to a vector that contains information about the entire source sentence. This means that the 'knowledge source' of NMT systems is the encoding vector which is supposed to contain all the necessary information to correctly translate the source into a target sentence. In this study, we want to examine how SMT and NMT, relying on different 'knowledge sources' (phrase-tables vs encoding-vectors), deal with a covert semantic category on the sentential level such as 'aspect'. For SMT, we know that its knowledge source is in theory, insufficient. The phrase-tables cannot always cover all the necessary contextual information in order to determine the correct aspectual value of a sentence. The phrase-tables can, however, reflect something about the lexical aspect of the verbs it contains. The probabilities stored in the phrase-tables of an English-Spanish or English-French translation system should be able to reflect whether the verb has a strong preference for an imperfective or perfective tense (or not). In order to verify whether this is the case, we compiled a list of 206 English verbs. The verbs were classified into their prototypical or basic aspectual class (following Wilmet's taxonomy (Wilmet 1997)). We then verified whether the phrase-tables reflect the connection between the aspectual classes and the grammatical aspect expressed by the French/Spanish tenses. Unlike SMT, NMT does encode the entire sentence at once and should theoretically have sufficient[4] information at their disposition to decode the sentence correctly. However, NMT's encoding vectors that consist of a large number of real-valued numbers are very hard to understand and interpret. Inspired by the work of Shi et al (2016), we used a logistic regression model trained on the encoding vectors to verify whether the encodings contain aspectual information.

Once we have a better understanding of what aspectual knowledge SMT and NMT decoders have at their disposal, we aim to see how this is reflected in the actual translations. We hypothesize that SMT performs well in the prototypical cases where the grammatical aspect reflects the lexical aspect of a verb. However, it will fail to render the correct grammatical aspect for the more complex cases where other contextual elements (adverbial phrases, prepositional phrases etc.) change the aspectual meaning of the verb. For NMT, formulating a hypothesis is more complicated. We know that the source sentence vectors encode the entire sentence but we do not know whether they implicitly store any aspectual information. Furthermore, even if they do contain aspectual information, it might still be lost during the decoding process.

The remainder of this paper is structured as follows: Section 2 discusses related work on 'aspect' in MT; Section 3 explains in more detail the set-up of our experiments for SMT and NMT as well as their outcomes; finally, Section 4 presents the conclusions we draw from our work together with some experiments we have in mind for the future.

## 2. Related Work

Although tense, aspect and mood (often referred to as 'TAM') have received a lot of attention in linguistic fields such as formal semantics and logic (Richards 1982, McCawley 1971), within the field of data-driven MT (SMT and NMT) there has been little research on tense, aspect and mood. However, in knowledge-based MT systems such as Eurotra and Rosetta, comprehensive and detailed modules were included in order to deal with TAM-related translation difficulties (Van Eynde et al. 1985, Van Eynde 1988, Appelo 1986). Van Eynde (1985, 1988) integrates tense information by mapping the analysis of tense and aspect onto their meanings. Since many meanings can be assigned to one form, the assignment of meaning is followed by a disambiguation step where context factors (e.g. temporal adverbials and the Aktionsart of the basic proposition) are taken into account. Although Eurotra is a transfer-based RBMT system, the representations of tense and aspect are

4. When translating paragraphs, information of the previous and/or following sentences might be required in order to determine the correct tense. Hardmeier, Nivre and Tiedemann (2012) developed a decoding algorithm that can handle cross-sentence features. A detailed discussion and overview of discourse phenomena can be found in the thesis of Hardmeier (2014).

interlingual, which implies that their meaning can simply be copied during the actual transfer. Appelo (1986) intends to solve the problem of translating temporal expressions in natural languages within the Rosetta MT system framework. The Rosetta MT system uses the 'isomorphic grammar' method which attunes the grammars of languages such that "a sentence $s$ is a translation equivalent of a sentence *s'* if $s$ and *s'* have similar derivational histories" (Appelo 1986, p.1).

Within the field of data-driven SMT, the works focusing on specific problems related to tense and aspect remain scarce. There has been some research on tense prediction: Ye and Zhang (2005) focus on unbalanced levels of temporal reference between Chinese and English. They build a tense classifier in order to predict the English tense given a Chinese verb based on several lexical and syntactic features. Their tense classifier was trained on manually labeled data. However, they did not build their tense classifier into an MT system. Similarly Gong et al. (2012) focuses on tense prediction. They used an $n$-gram-based tense model in order to predict the appropriate tense. Focusing particularly on the English-French language pair, Loaiciga et al. (2014) developed a method for the alignment of verbs phrases by using GIZA++ (Och and Ney 2003), a POS tagger, a parser and several heuristics. They labeled the VPs with their tense and voice on both sides of the parallel text. Once the VPs are aligned and labeled, a tense predictor is trained on the labeled data based on several features.

Only a handful of research studies focused on particular problems related to 'aspect' in MT: Ye et al. (2007) report on a study of aspect marker generation for the English-Chinese language pair. They train an aspect marker classifier based on a maximum entropy model and achieve an overall classification accuracy of 78%. Meyer et al (2013) worked on the disambiguation of the *passé composé* and the *imparfait* when translating from English into French by focusing on narrativity. Their Maximum Entropy classifier was trained on data that was manually annotated for narrativity.

To the best of our knowledge, to date, no research has been done on tense and aspect in NMT. However, in order to reveal how much aspectual information is stored in NMT encoding vectors, we were inspired by the work of Shi et al. (2016). Their work uses the high-dimensional encoding vectors of a sequence to sequence model (*seq2seq*) to predict sentence-level labels. By training a logistic regression model on a set of labeled encoding vectors, they show that local and global syntactic information is contained within these vectors, but other types of syntactic information is still missing (e.g. subtleties such as attachment errors). Their logistic regression model to predict the voice (active/passive) using encoding cell states achieved an accuracy of 92%. Voice is however an overt phenomenon in English expressed by specific verbal constructions. It could be, therefore, that the model does not 'learn' but just preserves information about the word forms in the encoding vector. To learn more about aspect in NMT we perform a similar experiment with a more covert phenomenon (aspect) that can manifest itself in many different ways (verbal aspect, verb structure, adverbial phrases etc.).

## 3. Experiments

### 3.1 SMT

#### 3.1.1 Compilation of Verbs

We compiled a list of 206 English verbs from linguistic sources and classified them by their 'basic' lexical aspect according to Wilmet's taxonomy (Wilmet 1997).[5] According to Wilmet's taxonomy there are three basic lexical aspectual classes: *stative*, *dynamic telic* or *dynamic atelic*. Classifying verbs into their basic lexical class is not always straightforward. One could argue that a verb like

5. We opted for Wilmet's taxonomy instead of for the more well-known classes of Vendler (1957) (who distinguishes between 4 different aspectual verb classes: *states*, *activities*, *achievements* and *accomplishments*). Wilmet groups together Vendler's *achievements* and *accomplishments* based on their 'telic' feature. We believe this makes Wilmet's classes more generalizable and thus more appropriate for our purpose.

e.g. ‘to run’ or ‘to drive’ (and many more) can be both telic or atelic dynamic verbs as illustrated in Example 3:

**Example 3**

| | | |
|---|---|---|
| *(a)* | *“I ran.”* | *ATELIC* |
| *(b)* | *“I ran a mile.”* | *TELIC* |

However, our classification is based on a verb’s occurrence in its most basic proposition. We classify a verb as stative when it does not undergo any changes in between its initial and final stage. When a change does occur, as in Example 3, we classify the verb as dynamic. Since the verb ‘to run’ does not denote an inherent end-point in its most basic proposition ((3)a), we classify ‘to run’ as a dynamic atelic verb.

#### 3.1.2 Description of SMT system

The SMT systems are built with the Moses toolkit (Koehn et al. 2007). The data is tokenized and lowercased using the Moses tokenizer and lowercaser. Sentences longer than 60 tokens are filtered out. For training, we use the default Moses settings. We trained three systems on 1 million parallel sentences of: (1) the Europarl corpus (Koehn 2005), (2) the News Commentary corpus (Tiedemann 2012) and (3) The Open Subtitles corpus (Tiedemann 2009) for two language pairs: English–French and English–Spanish.

#### 3.1.3 Extracting Information from Phrase-Tables

The core component of phrase-based translation models and the main knowledge source for the decoder are the phrase-tables, which contain the probabilities of translating a word (or a sequence of words) from one language into another. These word sequences are referred to as *phrases*. Unlike the linguistic use of the term phrase (verb phrase, nouns phrase etc.), phrases in the phrase-tables do not have any linguistic motivation but are just an arbitrary sequence of words (Koehn 2009). All the knowledge that phrase-tables contain is extracted from the word and phrase alignments obtained from the parallel data they were trained on. Example 4 shows phrase-translations extracted from a phrase-table trained on Europarl data:

**Example 4**

<table>
<tr><th></th><th></th><th></th><th></th><th>p(en |fr)</th><th>lex(en |fr)</th><th>p(fr |en)</th><th>lex(fr |en)</th><th></th></tr>
<tr><td>worked</td><td>|||</td><td>ont abattu</td><td>|||</td><td>0.2</td><td>0.0078406</td><td>0.00085034</td><td>4.06111e-05</td><td>|||</td></tr>
<tr><td>worked</td><td>|||</td><td>ont collaboré</td><td>|||</td><td>0.0952381</td><td>0.13217</td><td>0.00340136</td><td>0.0013808</td><td>|||</td></tr>
<tr><td>worked</td><td>|||</td><td>ont fonctionné ,[6]</td><td>|||</td><td>1</td><td>0.211969</td><td>0.00085034</td><td>0.000345333</td><td>|||</td></tr>
<tr><td>worked</td><td>|||</td><td>ont travaillé</td><td>|||</td><td>0.037037</td><td>0.183833</td><td>0.00085034</td><td>4.57342e-05</td><td>|||</td></tr>
<tr><td>worked</td><td>|||</td><td>travaillait</td><td>|||</td><td>0.275862</td><td>0.148649</td><td>0.00680272</td><td>0.004663</td><td>|||</td></tr>
<tr><td>worked</td><td>|||</td><td>travaillant</td><td>|||</td><td>0.00671141</td><td>0.0110865</td><td>0.00170068</td><td>0.0021195</td><td>|||</td></tr>
</table>

As can be seen in Example 4, the possible phrase translations are followed by four scores: the inverse phrase translation probability (p(english|french)), the inverse lexical weighting (lex(english|french)), the direct phrase translation probability (p(french|english)) and the direct lexical weighting (lex(french |english)). We are interested in the probability of the French word (or phrase) given an English word, i.e. the direct phrase translation probability (p(french|english)). We extracted all *imparfait* translations of the English verbs and collected their probabilities.

6. Although this is a relatively ‘clean’ fraction of a phrase-table, phrase-tables do not only include a lot of lexical variation (‘abbatu’ vs ‘travaillé’) and morphological variation (‘travaillait’ vs ‘travaillant’) but also often include function words (‘the’ , ‘an’ , ...) and noise (‘,’,‘it’). [`http://www.statmt.org/book/slides/05-phrase-based-models.pdf`].

We first tagged the phrases with the POS-tagger provided by the python package `polyglot`.[7] The pre-trained POS-tagger recognises the 17 universal parts of speech for several languages including Dutch, French and Spanish (Castilian).[8] Afterwards, we identified the perfective tenses by searching for phrases that contain a conjugated present tense auxiliary verb 'to have' ('avoir', 'haber' or 'hebben' for French, Spanish and Dutch, respectively) followed by a past participle.[9] To cover the exceptions, for the French pronominal verbs as well as for 14 other verbs (and their derivatives)[10], the auxiliary verb 'être' ('to be') was used. The imperfective tense was identified by extracting only those verbs with particular endings[11] characteristic of the imperfective tense of the language in question. We included the irregular conjugations that are not covered by the general endings[12]. Since some verbs in the present indicative, the present subjunctive and the present conditional tense in Spanish, French and Dutch have endings that overlap with those of the imperfective tenses, we cleaned the extracted phrases afterwards and added the false positives to a list in order to remove them from our results.

By then dividing both the added *passé composé* and *imparfait* values by the total (*imparfait* and *passé composé*), we normalize the probabilities and obtain the probability of the *passé composé* and/or *imparfait* given the total *passé composé* and *imparfait* translations of a specific verb. Table 1 below illustrates the different tense probabilities of the verbs 'promised', 'hit', 'saw' and 'thought' based on the information extracted from SMT phrase-tables trained on 1M sentences of Europarl, the News Commentary Corpus and the OpenSubtitles Corpus for English–French:

| % Verbs | Imparfait | | | Passé composé | | |
|---|---|---|---|---|---|---|
| | NEWS | EUROPARL | OpenSubs | NEWS | EUROPARL | OpenSubs |
| promised | 20.63 | 6.12 | 1.07 | **79.37** | **93.88** | **98.93** |
| hit | 7.14 | 0.00 | 3.31 | **92.86** | **100.00** | **96.69** |
| said | 16.98 | 11.47 | 9.53 | **83.02** | **88.53** | **90.47** |
| thought | **74.32** | **72.78** | **62.70** | 25.68 | 27.22 | 37.30 |

Table 1: *Imparfait* and *passé composé* percentages for the verbs *promised*, *hit*, *saw* and *thought*.

As can be seen in Table 1, these four English verbs ('promised', 'hit', 'said' and 'thought') have a clear preference for one particular tense in French according to the information extracted from the phrase-tables. Except for the verb 'thought', which has a stative lexical aspect, the verbs 'hit' (telic dynamic), 'promised' (telic dynamic) and 'said' (atelic dynamic) are more commonly translated into *passé composé* than into *imparfait*. This is the case in all three corpora. However, we do also observe some differences across the corpora when looking at these particular verbs, e.g. in the NEWS domain the verbs are translated more often into an imperfective tense compared to the Europarl and OpenSubtitles corpora. These differences can be explained by the fact that the Europarl and OpenSubtitles corpora often contain reported speech and the perfective tense (*passé composé*) is linked more closely to the spoken domain (Labelle 1987, Grisot and Cartoni 2012).[13] We ought to note that we are working with parallel corpora and the data contained within such corpora might

7. http://polyglot.readthedocs.io/en/latest/
8. http://polyglot.readthedocs.io/en/latest/POS.html
9. We allowed for insertion of adverbs and/or pronouns in between the auxiliary and the participle.
10. Commonly known as the verbs of 'La Maison d'être': 'aller','(r)entrer','passer','(re)monter','(re)tomber','arriver','(re)naître','(re)descen 
'(res)sortir','retourner','rester','(re)partir','mourir'
11. French: '-ais','-ait','-ions','-iez','-aient'; Spanish:'-aba','-abas','-ábamos','-abais','-aban','-a','-amos','-as','-ais','-an' Dutch: '-te(n)','-de(n)'.
12. Verbs such as 'ir','ser' and 'ver' (Spanish) and the so-called 'strong' Dutch verbs.
13. This illustrates as well how the use of particular tenses can depend not only on their immediate context but also on a broader context, paragraph or even on an entire domain.

be susceptible to well-studied phenomena in human translation such as interference of the source on the target (Xiao 2015).[14]

Together with the information about the French past tenses the verbs can be translated into, we also extract all the possible *imparfait* and *passé composé* translations of the verb in order to, later on during our research, use the translations to label English source sentences according to the grammatical aspect expressed by their corresponding reference target sentences. By storing all verbs with their translations, we are able to automatically label sentences and evaluate the outputs of our translation systems. An example of the possible translations extracted from the phrase-tables for one verb is given in Example 5:

**Example 5**

> ***Verb:*** *"felt"*
>
> ***Passé Comp.:*** *a compris; a; considéré; a cru; a estimé; a jugé; a paru; a pensé; a ressenti; a semblé; a senti; a tenu; a trouvé; ai considéré; ai estimé; ai jugées; ai notées; ai pensé; ai ressenti; ai senti; ai trouvé; avons considéré; avons estimé; avons jugé; avons pensé; avons ressenti; ont considéré; ont estimé; ont pensé; ont ressenti; ont senti; ont trouvé; s' est senti; s' est sentie; s' est vue*
>
> ***Imparfait :*** *craignaient; estimaient; estimait; jugeait; paraissait; pensaient*

#### 3.1.4 Aspectual information in SMT

To arrive at a more general idea we further analyzed how the lexical aspect of English verbs correlates with the grammatical aspect expressed by the French (Table 1) and Spanish (Table 2) tenses. Based on the information extracted from the Europarl, OpenSubs and News corpora, we analyzed which tense appeared overall more often when translating from an English *simple past/present perfect*. We included the results for all verbs that had a 'strong' (>67%) preference for the perfective tense (*passé composé*) or the imperfective tense (*imparfait*).

As can be seen in Table 2, the French *passé composé* appears more often than the *imparfait* in all corpora. Furthermore, with respect to the lexical aspect classes assigned to the English verbs, we see some clear patterns. *Passé composé* is the most used verb tense when translating an English *simple past* or *present perfect* verb. It also appears to be the more flexible one: both stative and dynamic (telic and atelic) verbs can have a preference for *passé composé* in French. These finding are in accordance with the literature stating that the perfective viewpoint is dominant in French (Smith 2013). The results correspond as well with the results of a previous contrastive linguistics study by Grisot and Cartoni (2012) based on a corpus containing texts belonging to different domains (journalistic, judicial, literary and administrative). Grisot and Cartoni (2012) calculated the frequencies of the French tenses given the English *simple past* and *present perfect* and, in our experiments, in both the News and Europarl corpora the *passé composé* dominated, especially when translating from an English *present perfect* verb. For the English–French data, we also observe that the judicial and spoken domains (Europarl and OpenSubtitles) contain more verbs that have a preference for the perfective tense compared to the News domain.

While the usage of the *passé composé* extends over different lexical aspect classes, the *imparfait* has a clear preference for stative verbs. English atelic dynamic verbs can also be translated into a French *imparfait*, however, our analysis revealed that none of the telic dynamic verbs were translated as an *imparfait*. Their telicity is difficult to unite with the imperfectivity of the French *imparfait*.

From the analysis above, performed on data extracted from SMT phrase-tables, we conclude that SMT's knowledge source indirectly possesses information about the lexical aspect of verbs.

14. Furthermore, the original language of the data can be either of the two languages involved, i.e. when working with an English-French corpus, the original data could have been English that was translated into French, or vice versa.

Furthermore, there is indeed a relation between the English lexical aspect assigned to verbs and the grammatical aspect of the French tense they are translated into.

| ENGLISH - FRENCH | SIMPLE PAST | | | PRESENT PERFECT | | |
|---|---|---|---|---|---|---|
| | EUROPARL | OPENSUBS | NEWS | EUROPARL | OPENSUBS | NEWS |
| PASSÉ C. | 73% | 73% | 63% | 98% | 99% | 98% |
| stative | 11% | 10% | 9% | 22% | 15% | 20% |
| atelic | 49% | 45% | 42% | 45% | 45% | 41% |
| telic | 40% | 45% | 49% | 33% | 39% | 40% |
| IMPARFAIT | 27% | 27% | 37% | 2% | 1% | 2% |
| stative | 93% | 84% | 84% | 75% | 100% | 100% |
| atelic | 7% | 16% | 16% | 0% | 0% | 0% |
| telic | 0% | 0% | 0% | 25% | 0% | 0% |

Table 2: English lexical verb classes VS grammatical aspect of French tenses.

We performed a similar analysis on the information extracted from the English–Spanish phrase-tables trained on the same corpora (Europarl, News, and OpenSubtitles) producing comparable results. These results are summarized in Table 3. As in French, the Spanish past tenses *pretérito indefinido* and *pretérito imperfecto* are linked with different grammatical aspect. Unlike French, where the *passé composé* became an equivalent (and almost completely replaced) the *simple past* tense, the *pretérito indefinido* is still widely used. Although Spanish has a past tense that is formally very similar to the *passé composé* (*pretérito perfecto*), the perfective aspect is marked by the *pretérito indefinido* while the imperfective aspect is marked by the *pretérito imperfecto*.

Looking at Table 3, it results that the Spanish *pretérito indefinido* and *pretérito imperfecto* behave similarly to the French past tenses. As in the French data, the perfective tense is dominant in all corpora for both *simple past* and *present perfect* verbs. The perfective *preterite* past tense occurs with verbs from all lexical aspect classes while the imperfect past tense appears only as a translation of stative and atelic dynamic verbs, with a clear preference for stative verbs.

Both the English–French and English–Spanish tables showed the dominance of the perfective tense over the imperfective one as well as the limited use of imperfective tenses with respect to certain verb classes (specifically the telic dynamic verbs). Given the fact that telic verbs present a completed action and the *imparfait* and *imperfecto* have an imperfective aspect presenting something ongoing, the lexical and grammatical aspect do not match, which explains why we had no occurrences in our data of telic dynamic verbs being translated into a tense with an imperfective aspect. The fact that telic verbs in Romance languages do not combine well with an imperfective tense was also noted by King and Suñer (1980).

| ENGLISH - SPANISH | SIMPLE PAST | | | PRESENT PERFECT | | |
|---|---|---|---|---|---|---|
| | EUROPARL | OPENSUBS | NEWS | EUROPARL | OPENSUBS | NEWS |
| PRET. | 75% | 71% | 74% | 81% | 97% | 97% |
| stative | 12% | 11% | 15% | 14% | 9% | 15% |
| atelic | 44% | 43% | 42% | 39% | 44% | 41% |
| telic | 45% | 46% | 44% | 47% | 47% | 44% |
| IMPERF. | 25% | 29% | 26% | 19% | 3% | 3% |
| stative | 82% | 83% | 84% | 83% | 0% | 50% |
| atelic | 18% | 17% | 16% | 17% | 100% | 50% |
| telic | 0% | 0% | 0% | 0% | 0% | 0% |

Table 3: English Lexical verb classes vs grammatical aspect of Spanish tenses.

In order to verify our method, we performed the same experiment with English–Dutch data.[15] English and Dutch have a similar verb system with tenses that are inflected similarly. There are some cases where translating the *simple past* or *present perfect* tense from English into Dutch might result in a negative transfer, but, Dutch past tenses (unlike the French and Spanish ones) do not grammaticalize aspect. The results shown in Table 4 confirm this. There is almost a one-to-one correspondence between the English *simple past* and the Dutch *onvoltooid verleden tijd* (OVT): In Europarl, 90% of the English simple past verbs are translated by a Dutch OVT, in the OpenSubtitles 97%. Similarly, the English *present perfect* 'matches' the Dutch *voltooid tegenwoordige tijd* (VTT) with 97% (Europarl) and 95% (OpenSubtitles).

| ENGLISH - | SIMPLE PAST | | PRESENT PERFECT | |
|---|---|---|---|---|
| DUTCH | EUROPARL | OPENSUBS | EUROPARL | OPENSUBS |
| OVT | **90%** | **97%** | 3% | 5% |
| stative | **42%** | **27%** | 60% | 25% |
| atelic | **36%** | **46%** | 40% | 50% |
| telic | **22%** | **27%** | 0% | 25% |
| VTT | 10% | 3% | **97%** | **95%** |
| stative | 0%[16] | 0% [16] | **19%** | **17%** |
| atelic | 70% | 0% | **43%** | **42%** |
| telic | 30% | 100% | **38%** | **41%** |

Table 4: English Lexical verb classes vs Dutch tenses.

**Intermediate Conclusion** According to Vinay (1995) and Filip (2012), aspect should be regarded as a non-lexical property that cannot be assigned to separate words but constitutes a property of an entire sentence. Although we do not disagree with this view, we do believe (and see this confirmed by the results in Section 3.1.4) that the lexical aspect of a verb often correlates with the grammatical aspect expressed by Spanish and French past tenses. Some verbs (especially the stative and telic activity verbs) do show a clear preference for a perfective or imperfective tense.

Nonetheless, contextual triggers (e.g. adverbs/adverbial phrases) can influence and change the aspectual 'value' of a sentence and thus cause the verb to be translated into the tense with the opposite aspect of its lexical aspect. In the sentences in Example 6[17], the telic verb 'to lose' is translated in a 'neutral' context into the French perfective past tense. In Example 6, sentence (b), the same verb is translated into the French imperfective past tense due to the aspectual meaning of the adverbial phrase 'all the time'. In this case, in an PB-SMT system an $n$-gram of size 5 would be needed in order to have the verb 'lost' with the adverbial phrase 'all the time' appearing in one phrase. Even if the size of the $n$-grams is set to 5, you would need an exact match with the 5-gram 'lost Manny all the time' to be able[18] to retrieve that particular translation.

**Example 6**

| | | | |
|---|---|---|---|
| *(a)* | *"I lost."* | *"J'ai perdu."* | *TELIC* |
| *(b)* | *"I lost Manny all the time"* | *"Je perdais Manny tout le temps."* | *ATELIC* |

15. We used the exact same training data for our Dutch MT system as for our English–French and English–Spanish systems. However, the NEWS corpus is not available in Dutch which is why we do not provide data from the news domain.
16. As mentioned in Section 3.1.4, the tables (2-4) only present the verbs that had a 'strong' (>67%) preference for a particular tense. Although it is possible to translate a stative English simple past verb into a Dutch VTT, there were no stative English simple past verbs (0% in both Europarl and the OpenSubtitles corpora) that had a strong inclination for the VTT.
17. The example sentences (in Examples 6 and 7) were extracted from the English-French OpenSubtitles Corpus.
18. We want to stress the fact that you would be able to retrieve it but there is still no guarantee that that particular 5-gram would eventually be selected.

Nevertheless some verbs do not show such a clear preference. An atelic activity verb such as 'walked', 'run' or 'eat' can easily be translated into both the perfective and imperfective French and Spanish tenses (even without any contextual triggers). This is illustrated in Example 7[17].

**Example 7**

| | | | |
|---|---|---|---|
| *(a)* | *"He ran."* | *"Il courrait."* | *ATELIC* |
| *(b)* | *"He ran."* | *"Il a couru."* | *TELIC* |

The information extracted from the English–French and English–Spanish phrase-tables from different corpora is in accordance with our hypothesis that verbs (belonging to different lexical aspect classes) often have a preference for a tense connected to a specific grammatical class. However, SMT does not have any means to extract contextual aspectual triggers from the context except for the (limited) $n$-grams stored in the phrase-tables. Therefore, we hypothesize that SMT performs well in terms of selecting the correct past tense in French and Spanish for verbs that have a strong lexical aspect. However, verbs that do not have such a clear lexical aspect and that rely more on the context to select the correct past tense in French and Spanish are most likely to cause more difficulties for a baseline SMT system. The actual translations of our SMT system will be evaluated in Section 3.2.4.

## 3.2 NMT

We will start by describing the NMT systems and the data set we trained on. Afterwards, our logistic regression model will be described in more detail followed by a discussion of its results.

### 3.2.1 Description of the NMT system

We carried out experiment with an encoder-decoder NMT model trained with the toolkit `nematus` (Sennrich et al. 2017). Our model was trained with the following parameters: *vocabulary size:* 45000, *maximum sentence length:* 60, *vector dimension:* 1024, *word dimension:* 500, *learning optimizer:* `adadelta`. In order to by-pass the out-of-vocabulary (OOV) problem and reduce the number of dictionary entries we use word-segmentation with byte-pair encoding (BPE) (Sennrich et al. 2015). We ran the BPE with 89500 operations.

### 3.2.2 Aspectual information in NMT

NMT does store information about the entire sentence in its encoding vectors, unlike SMT. A recent work by Shi et al. (2016) uses the high-dimensional encoding vectors of a sequence to sequence model (*seq2seq*) to predict sentence-level labels. They show that much syntactic information is contained within these vectors, but, other types of syntactic information is still missing. They trained an NMT system on 110M tokens of bilingual (English–French) data. They created a set of 10K English sentences that were labeled for voice (active or passive) and converted them with their learned NMT encoder into their corresponding encoding vectors (1000-dimensions). A logistic regression model was then trained on 9K sentences to learn to predict voice and tense based on the English encoding vectors. They tested their logistic regression model on 1K sentences and achieved an accuracy of 92.8%. Our work on discovering how much aspectual information is contained within an NMT system is inspired by this work as we also used a logistic regression model to predict 'aspect' based on the encoding vectors. However, it also differs from their experiments in two ways:

**(1)** First, a different 'voice' or 'tense' in a sentence manifest themselves 'overtly' in the English source sentence, i.e., with a different verbal construction for passive and active voices and different verbal forms for all the English tenses. The passive will e.g. always be characterized by a +[be]-construction, such as in Example 8:

**Example 8**

| | | |
|---|---|---|
| *(a)* | *"I taught French (to ...)"* | *ACTIVE* |
| *(b)* | *"I was taught French (by ...)"* | *PASSIVE* |

This is not the case for aspect in English. The simple past in English is neutral with respect to aspect but, as illustrated before in Example 2 and 6, contextual information (e.g. adverbs, adverbial phrases) can carry aspectual meaning. Determining the aspect of a sentence, if at all possible[19], is a more complex task: aspectual meaning can be conveyed by words with different parts of speech and can furthermore be combined together in complex ways to create aspectual meaning of a verbal expression. We therefore hypothesize that predicting the aspect of sentences based on their NMT encoding vectors is a more complex task.

**(2)** Second, the work from Shi et al. (2016) shows that encoding vectors capture certain linguistic information. Their study, however, does not include any information on the actual effect of this on the translation. Is this information also 'decoded' correctly?

In the next Section 3.2.3, our first task is to examine whether aspectual information is captured by the encoder at all.

### 3.2.3 Logistic Regression Model

In order to train our logistic regression model on the NMT encoding vectors, we need labeled English data. Since our NMT model is specifically trained to translate from English into French/Spanish, and French/Spanish require the past tense to make a distinction between two different past tenses that are each associated with an opposite aspectual value, we labeled the English sentences based on the aspectual value of the tense in the French/Spanish reference translations. As explained in Section 3.1.3, we did not only extract information about the preference of specific verbs for the one aspectual tense or the other but also the specific translations of the verbs themselves. Since our NMT system is trained on 3M OpenSubtitles sentence pairs, we trained an SMT system on the same data and extracted the possible imperfective (*imparfait* and *pretérito imperfecto*) and perfective (*passé composé* and *pretérito indefinido*) translations of our verbs. We used a separate set of the OpenSubtitles corpus to extract sentences with on the one hand, verbs in the English simple past, and on the other hand a French/Spanish reference sentence containing an *imparfait/pretérito imperfecto* or *passé composé/pretérito indefinido* verb. Based on the appearance of either an imperfective tense or a perfective tense in the reference translations, we automatically labeled the corresponding English sentences and limited the length of the sentence to 10 tokens.[20]

We randomly selected 40K labeled sentences, and generated for every sentence their encoding vector with the NMT system described above. Next, we trained a logistic regression using the python machine learning toolkit `scikit learn`[21] with the default settings.

### 3.2.4 Results

To test our logistic regression model we compiled 4 test sets of increasing difficulty. Each of the test sets contains 2K sentences. The reason why we compiled 4 different test sets is because of the results we obtained and described in Section 3.1.3. Our results showed that some verbs have a very strong basic lexical aspect which links them to a particular tense in French ( 'surprised', 'jumped' (IMP: 0% and PC: 100%) or 'weighed', 'sounded' (IMP: 100% vs PC: 0%)). Other verbs can easily be associated with an imperfective or perfective aspectual value and thus rely more on other factors

19. Sometimes information from the broader context (other sentences/paragraphs) is needed in order to determine the aspect correctly.
20. We limited the sentence length since we wanted only one label per sentence (i.e. one main verb) in order to be able to train a logistic regression model on the sentence vectors and the imperfective/perfective labels.
21. http://scikit-learn.org/stable/

apart from their lexical aspect in order to disambiguate between the two tenses ('reigned' (67% vs 33%), 'lived' (44% vs 56%)). Since we assumed that especially the verbs that do not have a strong lexical aspect (and thus no strong preference for a particular tense) are harder to translate, we created 4 test sets. The first test set is the 'general' one that contains all types of verbs, while the second test set does not include verbs that have more than 80% or less than 20% preference for a particular tense. It thus only contains verbs whose preference for either tense is between 20%-80%. Similarly, test set 3 and test set 4 contain verbs with a 30%-70%, and 40%-60% preference respectively, i.e. the more 'ambiguous' verbs in terms of aspect. We compared the predictions of our logistic regression model with the reference labels for the 4 test sets. To check our logistic regression model, we also computed a naive baseline performance, which represents the highest accuracy that would be obtained if all predictions would consist of only either 0's or 1's.[22] The results of the logistic regression model for French and Spanish can be found in Table 5.

| | | **100–0** | **80–20** | **70–30** | **60–40** |
|---|---|---|---|---|---|
| French | **LogReg** | 90.95% | 86.10% | 86.20% | 77.10% |
| | **Baseline** | 64.55% | 64.60% | 74.55 | 72.60% |
| Spanish | **LogReg** | 87.05% | 73.95% | 65.10% | 66.10% |
| | **Baseline** | 76.70% | 52.40% | 60.65% | 60.70% |

Table 5: Prediction accuracy of Logistic Regression Model on the French and Spanish vectors.

The results of Table 5 confirm our hypothesis that the more ambiguous verbs in terms of aspect are harder to predict for the logistic regression model since the prediction accuracy lowers over the test sets. In the general test set (referred to as "100–0" in Table 5) the accuracy of the logistic regression model is 90.95%. The accuracy drops when excluding verbs with a strong lexical aspect (test set 80–20 and 70–30) to 86.10% and 86.20%, respectively. The lowest score prediction accuracy is 77.10% for the fourth test set (60–40), containing verbs that, without any further context, are almost equally likely to be translated into either of both French past tenses.

The fact that a logistic model can extract certain aspectual information from the NMT encoding vectors does not guarantee that the decoder is able to. Therefore, in the next section we will examine in more detail the actual outputs of our NMT/SMT systems on the same 4 test sets.

### 3.3 Aspect in NMT/SMT Translations

A logistic regression model trained to label English source vectors with a particular tense achieved an accuracy of 90.95% (English–French) and 87.05% (English–Spanish), which are promising results. So far, however, we have not yet looked at the actual outputs of both systems. Therefore, we will now examine how the SMT and NMT translations compare (in terms of selecting an imperfective or perfective French/Spanish tense) with respect to the reference translations.

We translated the 4 test sets described in Section 3.2.4 with a NMT and SMT system trained on the same 3M OpenSubtitles sentences. In Section 3.1.3 we explained that, together with the *perfective* and *imperfective* preference of verbs, we also extracted all the translations stored in the SMT phrase-tables. As Example 5 in Section 3.1.3 shows, these translations do not contain all possible correct translations (often only the third person of verbs is represented e.g. 'felt' -'craignaient', 'estimaient', 'jugeait' etc.). Therefore, we first of all made sure our translations included all possible forms (in terms of persons) of the translations extracted. We also included the 'female' and 'plural' forms of the French participle. One such translation expansion is partially shown in Example 9:

22. The baseline represents the ratio of perfective to imperfective labels.

**Example 9**

***Verb:*** *"felt"*

***Passé Comp.:*** *ai compris; ai comprise; ai comprises; as compris; as comprise; as comprises;* a compris*; a comprise; a comprises; avons compris; avons comprise; avons comprises; avez compris; avez comprise; avez comprises; ont compris; ont comprise; ont comprises;* ai consideré*; ai considerée; ai considerées; as consideré; ...*

***Imparfait:*** *craignais; craignait, craignions; craigniez;* craignaient*; estimais;* estimait*; estimions, estimiez;* estimaient*; ...*

Based on the verbs in the source sentence and their translations, we were able to automatically evaluate the outputs of our translation systems. The results of our translated test sets for French are presented in Table 6 and for Spanish are presented in Table 7. We observe, as we did with the logistic regression model, that also for the translations, our test sets present different difficulty levels. Both NMT and SMT perform best on the data containing all types of verbs (test set 1 *100–0*). Performance decreases over the other three test sets with +/- 12% for both NMT and SMT. The logistic regression model indicated we can accurately (90.95%) predict the correct grammatical aspect for the French tense in the general test set based on the vector encodings. We do not see this same accuracy reflected in the actual translations of the general test set (79%). This implies that part of the aspectual information is lost during decoding.

| | | 100–0 | | 80–20 | | 70–30 | | 60–40 | |
|---|---|---|---|---|---|---|---|---|---|
| Ref. | Trans. | SMT | NMT | SMT | NMT | SMT | NMT | SMT | NMT |
| IMP | PC | 114 | 103 | 214 | 221 | 246 | 229 | 248 | 170 |
| PC | IMP | 26 | 40 | 72 | 93 | 86 | 103 | 93 | 171 |
| Correct (in %) | | 77.90% | 79.05% | 72.50% | 71.55% | 72.05% | 70.85% | 65.55% | 67.20% |

Table 6: Translation accuracy SMT vs NMT for the OpenSubtitles test sets for the English-French language pair.

Surprisingly, the performance of SMT and NMT is very comparable on the all test sets although their knowledge sources are different. This indicates that the lexical aspect of a verb plays an important role when selecting the tense with the correct grammatical aspect. This statement is consistent with the observations in Ye et al (2007) and Olsen et al (2001). Ye et al (2005, 2006, 2007) reported on the high utility of lexical aspectual features in selecting a tense. Similarly, Olsen et al (2001) reported on the significance of the telicity of verbs in order to reconstruct the tense for Chinese-to-English translation.

For English–Spanish (Table 7), we obtained similar results. However, the results are overall lower than the ones obtained for the English–French systems. The tenses in the translations only correspond to the tense of the reference translations in 46.50% [23] of cases for SMT, and 57.70% for NMT on the general test set (100–0). We analyzed some of the translations in order to identify why the results are lower than for the English–French systems and saw that often, our NMT and SMT systems opted for another Spanish tense: the *pretérito perfecto compuesto*. In the future, we would like to further extend our work in order to cover this additional tense. Unlike the English–French results, the English–Spanish NMT systems consistently outperform the SMT systems in terms of selecting the same tense as the reference.

We see, for both English–French and English–Spanish language pairs that the tense prediction of the logistic regression model is more accurate than the tenses in the NMT outputs. This is

23. Since the perfective tense appears more in the test sets than the imperfective one and SMT regularly opts for an imperfective tense (while the reference is perfective), percentages can be below 50%.

| | | 100–0 | | 80–20 | | 70–30 | | 60–40 | |
|---|---|---|---|---|---|---|---|---|---|
| Ref. | Trans. | SMT | NMT | SMT | NMT | SMT | NMT | SMT | NMT |
| IMP | PC | 71 | 84 | 191 | 209 | 365 | 358 | 208 | 169 |
| PC | IMP | 98 | 84 | 225 | 178 | 251 | 229 | 361 | 348 |
| Correct (in %) | | 46.50%[23] | 57.70% | 58.25% | 62.45% | 49.60%[23] | 51.95% | 46.70%[23] | 49.15%[23] |

Table 7: Translation accuracy SMT vs NMT for the OpenSubtitles test sets for the English-Spanish language pair.

most likely due to the fact that the logistic regression model is trained for one specific task while the decoder has to take care of multiple tasks simultaneously such as word translations, word reordering, etc.

## 4. Conclusions and Future Work

We investigated what kind of aspectual information SMT and NMT could grasp and how this is reflected in their translations. For SMT, we saw the lexical aspect of verbs reflected in their 'knowledge source', i.e. the phrase-tables. SMT's knowledge is limited to the size of the $n$-grams in the phrase-tables, so they cannot cover other aspectual factors that appear in a sentence (in case they fall out of the $n$-gram range). We hypothesized this would be particularly problematic for those verbs that do not have a 'strong' lexical aspect, which we saw confirmed by our results.

NMT does have the means to store information about the entire source sentence. By using a logistic regression model trained on the encoding vectors, we discovered that NMT encoding vectors do capture aspectual information. The evaluation of the actual outputs of the NMT and SMT systems in terms of *imperfective* or *perfective* tense choice revealed that NMT and SMT perform very similarly on all test sets. Although aspect can accurately (90.95%) be predicted from the encoding vectors by a logistic regression model, the NMT decoder loses some of this information during the decoding process.

In future work, we would like to manually evaluate part of the translations to confirm whether they confirm our automatic evaluations results. Furthermore, we will also investigate how to explicitly utilize the aspectual information in the decoding process to improve translation performance. The initial idea is to use a Convolutional Neural Network (CNN) to learn the aspect features and then explicitly integrate them into the emission of target words as probabilistic conditions, or use a selective attention mechanism that can selectively learn some special linguistic information such as aspectual knowledge for a target word generation focusing on the verb.

## 5. Acknowledgements

This work has been supported by the Dublin City University Faculty of Engineering & Computing under the Daniel O'Hare Research Scholarship scheme and by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

## References

Appelo, Lisette (1986), A Compositional Approach to the Translation of Temporal Expressions in the Rosetta System, *Proceedings of the 11th coference on Computational linguistics*, Association for Computational Linguistics, pp. 313–318.

Appelo, Lisette (1993), *Categorial Divergences in a Compositional Translation System*, PhD thesis, University of Utrecht, Utrecht, The Netherlands.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2015), Neural Machine Translation by Jointly Learning to Align and Translate, *Proceedings of International Conference on Learning Representations (ICLR2015)*, San Diego, USA.

Čech, Radek, Ján Mačutek, and Zdeněk Žabokrtskỳ (2011), The Role of Syntax in Complex Networks: Local and Global Importance of Verbs in a Syntactic Dependency Network, *Physica A: Statistical Mechanics and its Applications* **390** (20), pp. 3614–3623, Elsevier.

Cho, Kyunghyun, Bart van Merriënboer, Çalar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio (2014), Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, *Proceedings of Empirical Methods on Natural Language Processing (EMNLP 2014)*, Doha, Qatar.

Comrie, Bernard (1976), *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*, Vol. 2, Cambridge University Press, Cambride, UK.

Dagan, Ido, Alon Itai, and Ulrike Schwall (1991), Two Languages Are More Informative Than One, *Proceedings of the 29th annual meeting on Association for Computational Linguistics (ACL)*, Berkeley, California, USA, pp. 130–137.

Dowty, David R. (1986), The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics?, *Linguistics and philosophy* **9** (1), pp. 37–61, Springer, Berlin, Germany.

Engel, Dulcie M. (1998), A Perfect Piece? The Present Perfect and Passé Composé in Journalistic Texts, *Belgian Journal of Linguistics* **12** (1), pp. 129–147, John Benjamins Publishing Company, Amsterdam, The Netherlands.

Filip, Hana (2012), Lexical Aspect, *The Oxford Handbook of Tense and Aspect* pp. 721–751, Oxford University Press, Oxford, UK.

Fischer, Susan and Bonnie Gough (1978), Verbs in American Sign Language, *Sign Language Studies* **18** (1), pp. 17–48, Gallaudet University Press, Washington, USA.

Garey, Howard B (1957), Verbal Aspect in French, *Language* **33** (2), pp. 91–110, Linguistic Society of America, USA.

Gong, Zhengxian, Min Zhang, Chewlim Tan, and Guodong Zhou (2012), N-gram-based Tense Models for Statistical Machine Translation, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (ACL2012)*, Association for Computational Linguistics, pp. 276–285.

Grisot, Cristina and Bruno Cartoni (2012), Une Description Bilingue des Temps Verbaux: Étude Contrastive en Corpus, *Nouveaux Cahiers de Linguistique française* **30**, pp. 101–117.

Hardmeier, Christian (2014), *Discourse in Statistical Machine Translation*, PhD thesis, Acta Universitatis Upsaliensis, Uppsala, Sweden.

Hardmeier, Christian, Joakim Nivre, and Jörg Tiedemann (2012), Document-wide Decoding for Phrase-based Statistical Machine Translation, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics, pp. 1179–1190.

Hogeweg, Lotte, Helen De Hoop, and Andrej Malchukov (2009), *Cross-linguistic Semantics of Tense, Aspect, and Modality*, Vol. 148, John Benjamins Publishing, Amsterdam, The Netherlands.

King, Larry D and Margarita Suñer (1980), The Meaning of the Progressive in Spanish and Portuguese, *Bilingual Review/La Revista Bilingüe* **7** (3), pp. 222–238, JSTOR.

Koehn, Philipp (2005), Europarl: A Parallel Corpus for Statistical Machine Translation, *MT summit*, Vol. 5, Citeseer, pp. 79–86.

Koehn, Philipp (2009), *Statistical Machine Translation*, Cambridge University Press, Cambridge, UK.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Federico Marcello Callison-Burch, Chris, Cowan Brooke Shen Wade Moran Christine Zens Richard Dyer Chris Bojar Ondrey Constantin Alexandra Bertoldi, Nicola, and Evan Herbst (2007), Moses: Open-Source Toolkit for Statistical Machine Translation, *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions (ACL2007)*, Association for Computational Linguistics, pp. 177–180.

Labelle, Marie (1987), L'Utilisation des temps du passé dans les narrations françaises: Le Passé Composé, L'lmparfait et Le Présent Historique, *Revue romane*.

Loaiciga, Sharid, Thomas Meyer, and Andrei Popescu-Belis (2014), English-French Verb Phrase Alignment in Europarl for Tense Translation Modeling, *The Ninth Language Resources and Evaluation Conference*.

McCawley, James D (1971), Tense and Time Reference in English.

Meyer, Thomas, Cristina Grisot, and Andrei Popescu-Belis (2013), Detecting Narrativity to Improve English to French Translation of Simple Past Verbs, *Proceedings of the 1st DiscoMT Workshop at ACL 2013 (51st Annual Meeting of the Association for Computational Linguistics)*.

Miller, George A and Christiane Fellbaum (1991), Semantic networks of English, *Cognition* **41** (1), pp. 197–229, Elsevier.

Moens, Marc (1987), Tense, Aspect and Temporal Reference., The University of Edinburgh, Edinburgh, UK.

Och, Franz Josef and Hermann Ney (2003), A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics* **29** (1), pp. 19–51.

Olsen, Mari, David Traum, Van Ess-Dykema, Amy Weinberg, et al. (2001), Implicit Cues for Explicit Generation: Using Telicity as a Cue for Tense Structure in a Chinese to English MT System, *Technical report*.

Richards, Barry (1982), Tense, Aspect and Time Adverbials, *Linguistics and Philosophy* **5** (1), pp. 59–107, Springer.

Santos, Diana (2004), Translation-based Corpus Studies: Contrasting English and Portuguese Tense and Aspect systems., Rodopi, Amsterdam.

Sennrich, Rico, Barry Haddow, and Alexandra Birch (2015), Neural Machine Translation of Rare Words with Subword Units, *arXiv preprint arXiv:1508.07909*.

Sennrich, Rico, Orhan Firat, Kyunghyun Cho, Alexandra Birch, Barry Haddow, Julian Hitschler, Marcin Junczys-Dowmunt, Samuel L"aubli, Antonio Valerio Miceli Barone, Jozef Mokry, and Maria Nadejde (2017), Nematus: a toolkit for neural machine translation, *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for*

*Computational Linguistics*, Association for Computational Linguistics, Valencia, Spain, pp. 65–68.

Shi, Xing, Inkit Padhi, and Kevin Knight (2016), Does String-Based Neural MT Learn Source Syntax?, *Proceedings of Empirical Methods on Natural Language Processing (EMNLP 2016).*

Smith, Carlota S. (2013), *The Parameter of Aspect*, Vol. 43, Springer Science & Business Media, Dordrect, The Netherlands.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le (2014), Sequence to sequence learning with neural networks, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014*, Montreal, Quebec, Canada.

Tiedemann, Jörg (2009), News from OPUS - A collection of multilingual parallel corpora with tools and interfaces, *in* Nicolov, N., K. Bontcheva, G. Angelova, and R. Mitkov, editors, *Recent Advances in Natural Language Processing*, Vol. V, John Benjamins Publishing, Borovets, Bulgaria, pp. 237–248.

Tiedemann, Jörg (2012), Parallel data, tools and interfaces in OPUS, *in* Chair), Nicoletta Calzolari (Conference, Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), Istanbul, Turkey.

Van Eynde, Frank (1988), The Analysis of Tense and Aspect in Eurotra, *Proceedings of the 12th Conference on Computational Linguistics - Volume 2*, COLING '88, Budapest, Hungary, pp. 699–704.

Van Eynde, Frank, Louis des Tombe, and Fons Maes (1985), The Specification of Time Meaning for Machine Translation, *Proceedings of the Second Conference on European Chapter of the Association for Computational Linguistics*, EACL '85, Geneva, Switzerland, pp. 35–40.

Vanmassenhove, Eva, Jinhua Du, and Andy Way (2016), Improving Subject-Verb Agreement in SMT, Proceedings of the Fifth Workshop on Hybrid Approaches to Translation: HyTra (EAMT), Riga, Latvia.

Vendler, Zeno (1957), Verbs and Times, *The Philosophical Review* pp. 143–160, JSTOR.

Vinay, Jean-Paul and Jean Darbelnet (1995), *Comparative Stylistics of French and English: a Methodology for Translation*, Vol. 11, John Benjamins Publishing.

Wang, Longyue, Zhaopeng Tu, Xiaojun Zhang, Hang Li, Andy Way, and Qun Liu (2016), A Novel Approach to Dropped Pronoun Translation, *arXiv preprint arXiv:1604.06285.*

Wilmet, M. (1997), *Grammaire Critique du Français*, Duculot, France.

Xiao, Richard (2015), Source Language Interference in English-to-Chinese Translation, *Yearbook of Corpus Linguistics and Pragmatics 2015*, Springer, pp. 139–162.

Ye, Yang and Zhu Zhang (2005), Tense Tagging for Verbs in Cross-lingual Context: A Case Study, *International Conference on Natural Language Processing*, Springer, pp. 885–895.

Ye, Yang, Karl-Michael Schneider, and Steven Abney (2007), Aspect Marker Generation in English-to-Chinese Machine Translation, *Proceedings of MT Summit XI*, Citeseer, Copenhagen, Denmark.

Ye, Yang, Victoria Li Fossum, and Steven Abney (2006), Latent Features in Automatic Tense Translation between Chinese and English, *Proceedings of the 5th SIGHAN Workshop on Chinese Language Processing*, pp. 48–55.