If you've seen some, you've seen them all: Identifying variants of multiword expressions

Caroline Pasquer, Agata Savary, Carlos Ramisch, Jean-Yves Antoine

If you've seen some, you've seen them all: Identifying variants of multiword expressions

Caroline Pasquer
University of Tours
France

Agata Savary
University of Tours
France

Carlos Ramisch
Aix Marseille Univ,
Université de Toulon,
CNRS, LIS, Marseille, France

Jean-Yves Antoine
University of Tours
France

first.last@(univ-tours|lis-lab).fr

Abstract

Multiword expressions, especially verbal ones (VMWEs), show idiosyncratic variability, which is challenging for NLP applications, hence the need for VMWE identification. We focus on the task of variant identification, i.e. identifying variants of previously seen VMWEs, whatever their surface form. We model the problem as a classification task. Syntactic subtrees with previously seen combinations of lemmas are first extracted, and then classified on the basis of features relevant to morpho-syntactic variation of VMWEs. Feature values are both absolute, i.e. hold for a particular VMWE candidate, and relative, i.e. based on comparing a candidate with previously seen VMWEs. This approach outperforms a baseline by 4 percent points of F-measure on a French corpus.

Title and Abstract in French

Trait pour trait identiques ? Identification de variantes d'expressions polylexicales

Les expressions polylexicales (EP), et parmi elles plus particulièrement les EP verbales (EPV), se caractérisent par une grande variabilité idiosyncrasique de forme. La détection et l'identification de ces EPV variées pose ainsi un réel défi à la réalisation d'applications langagières robustes. Cet article met l'accent sur la tâche d'identification dans un corpus de variantes d'une EP verbale déjà rencontrées. Il propose une stratégie d'identification basée sur l'extraction de formes candidates à partir de patrons syntaxiques, suivie de leur classification basée sur des caractéristiques morphologiques et syntaxiques. Ces propriétés sont à la fois absolues (c.-à-d. concernent l'entité considérée) ou relatives (c.-à-d. issues de la comparaison avec des EPV déjà rencontrées). Les performances du système résultant ont été évaluées sur un corpus francophone. Elles montrent une amélioration de 4 points de F-mesure par rapport à une baseline bien établie.

1 Introduction

Multiword expressions (MWEs) such as ***red tape***, ***by and large***, *to* ***make*** *a* ***decision*** and *to* ***break*** *one's* ***heart*** are combinations of words exhibiting unexpected lexical, morphological, syntactic, semantic, pragmatic and/or statistical behavior (Baldwin and Kim, 2010). Most prominently, they are semantically non-compositional, that is, their meaning cannot be deduced from the meanings of their components and from their syntactic structure in a way deemed regular. For this reason, the presence of MWEs in texts calls for dedicated treatment, whose prerequisites include their automatic identification.

The goal of *MWE identification* is, given some input running text, to identify all MWEs' lexicalized components present in it (Schneider et al., 2016; Savary et al., 2017).[1] Such systems face three main

[1] A *lexicalized component* of a MWE, or *component* for short – marked in bold in examples – is one which is always realized by the same lexeme. For instance, the noun *decision* is lexicalized in *to* ***make*** *a* ***decision*** but the determiner *a* is not, because it can be replaced by other determiners (e.g. *to* ***make*** *this/any/the* ***decision***).

challenges: variability, ambiguity and discontinuity (Constant et al., 2017). Variability refers to the fact that the components of the same MWE can occur in different surface forms because of morphological and/or syntactic transformations, as in example (1) vs. (3) below. Variability is closely related to the issue of unseen data in machine learning – items whose surface forms differ between test and training data are usually harder to identify than those seen in identical forms (Augenstein et al., 2017). MWE ambiguity, conversely, makes the identification of seen MWEs harder, because a particular combination of words can be a MWE in some contexts, as in (1) and (2), but not necessarily all, because of *literal readings* (4) and *coincidental co-occurrence* of the MWE components (5) (Savary and Cordeiro, 2018).[2] Finally, external elements can occur in between MWEs' components, both seen and unseen (1)–(3). Discontinuity is known to be a challenge notably for methods based on sequence labeling.

(1) *He **broke** the **heart** of his class mate.*
(2) *He **broke** her young **heart** when he **let** her **down**.*
(3) *Just think of all the **hearts broken** by him.*
(4) *He broke her chocolate heart when he stepped on it.*
(5) *When her toy broke, her young heart was in sorrow.*

In this paper we address the three aforementioned challenges at a time by considering a sub-problem of MWE identification, namely *MWE variant identification*, that is, the identification of occurrences of known (seen) MWEs and their linguistic variants. For instance, suppose that we have seen the annotated MWE occurrence in (1) in some training data. We would like to correctly predict that the same MWE occurs also in (2) and (3) but not in (4) and (5). Notice that we do not aim at identifying the MWE ***let down*** in example (2) if it has not been previously seen. This focus on MWE variants has important theoretical and practical motivations discussed in Sec. 2.

MWEs are known to have specific variability profiles, e.g. the idiom in (1)–(3) admits passivization and noun inflection, while others do not, e.g. *he was **pulling** my **leg*** 'he was kidding me' vs. *he was pulling my legs*; *my leg was pulled*. Our research question is whether this MWE-specific variability profile may be captured automatically and leveraged for high-quality MWE variant identification. The idea is that defining a variability profile in terms of linguistically informed variation-oriented features should help, on the one hand, to bring different variants of the same MWE together and, on the other hand, to distinguish idiomatic occurrences of a MWE from its literal readings. We are specifically interested in morphological and syntactic variants of verbal MWEs (VMWEs). This phenomenon is particularly challenging in morphologically rich languages, notably due to the internal inflection of the VMWE components and to the relatively free word order. In this paper we are interested in VMWEs in French, which exhibits rich verbal inflection and moderate inflection of nouns and adjectives.

This paper is organized as follows. We discuss the state of the art in MWE variant modeling and identification (Sec. 2). We address aspects of VMWE variability in French (Sec. 3), and we describe the corpus used for our experiments (Sec. 4). Follows a presentation of our a baseline variant identification method (Sec. 5). We then propose linguistic features for classification and their extraction process, as well as the architecture of our VMWE variant identification system (Sec. 6). Finally, we present and analyze our results (Sec. 7), before we conclude and propose perspectives for future work (Sec. 10).

2 Related work

Variability, often referred to as flexibility, has been considered a key property of MWEs in various linguistic studies in the past. Gross (1988) analyzes French adjective-noun compounds and shows that their fixedness (the contrary of flexibility) is a matter of scale rather than a binary property. Tutin (2016) confirms this hypothesis with a corpus study of the 30 most frequent verb-noun expressions in French, and defines 6 variability levels. Nunberg et al. (1994) hypothesize a strong correlation between semantic decomposability and syntactic flexibility. They conjecture that MWEs such as *to **pull strings*** admit a large range of variants (e.g. *to **pull** political **strings**, the many **strings** he **pulled***) because their components *pull* and *strings* can be paraphrased by *use* and *influence*. This hypothesis has been criticized by Sheinfux et al. (2017), who stipulated that the degree of flexibility of Hebrew VMWEs depends on the links

[2] Literal readings and coincidental occurrences of MWEs are maked by wavy and dashed underlining, respectively.

between their literal and idiomatic readings rather than on their semantic decomposability. Inflectional and syntactic flexibility were also a major criterion for classifying MWEs into fixed, semi-fixed (subject to internal inflection only) and syntactically flexible expressions by Sag et al. (2002). This scale-wise variability challenges the traditional lexicon vs. grammar division of language modeling.

MWE variability has also been considered a major challenge in various NLP models and applications. Firstly, variants are pervasive. Jacquemin (2001) studies multiword nominal terms in French and English and shows that as many as 28% of their occurrences in texts correspond to variants of canonical forms listed in lexicons. Secondly, variants challenge automatic identification of MWEs, as discussed below. Finally, variant conflation is crucial for downstream applications such as information extraction (Savary and Jacquemin, 2003) and entity linking (Hachey et al., 2013).

Existing approaches to MWE identification partly model MWE variability to some extent. *Rule-based methods* often rely on MWE lexicons and a matching procedure. They sometimes extensively address MWE variation by morphological processors combined with rich finite-state patterns (Krstev et al., 2014) or unification meta-grammars (Jacquemin, 2001). Rule-based methods were often combined with statistical measures in the domain of multiword term extraction (Savary and Jacquemin, 2003), and explicitly addressed variation. But they can hardly distinguish literal from idiomatic readings, weakly cover discontinuous MWEs, and cannot generalize to MWEs absent from the lexicon and to new, initially uncovered, variation patterns. *Sense disambiguation methods* often focus on ambiguous known MWEs and neglect variant identification (Fazly et al., 2009). *Sequence taggers* learn identification models from annotated data based on regularities in sequences of tokens (Constant et al., 2013; Schneider et al., 2014). They are well suited for continuous seen MWEs and neutralize inflection when lemmas are available. They cope, however, badly with syntactic variation whenever it leads to discontinuities. *MWE-aware parsers* identify MWEs as a by-product of parsing a sentence (Green et al., 2013; Constant and Nivre, 2016). They often can deal with both morphological and syntactic variants, even though they are usually less accurate for highly discontinuous and variable MWEs, given that parsing is a hard problem by itself, and MWE-specific features increase data sparseness. *Parsing-based MWE identifiers* use generic parsers as providers of MWE candidates for classification (Vincze et al., 2013). They are less subject to data sparseness and can cope with discontinuities, provided that the underlying parser correctly handles long-distance dependencies. However, they may under- or overgenerate some syntactic transformations allowed by a MWE, e.g. indirect dependencies, as in Fig. 1.c and 1.f.

As a conclusion, it seems that MWE variant identification, although crucial both for corpus-based linguistic studies and for downstream NLP applications, has not yet received a satisfactory solution. Moreover, performances in solving this task are rarely explicitly reported on. Instead, the performances on seen vs. unseen data are addressed, and it is not always clear how these notions are understood. For instance, Constant et al. (2013) observe that 25% of the MWEs in their test corpus are unseen in the training data, and that at most 19% of them could be correctly identified by their system based on sequence labeling. However, the authors do not specify how unseen MWEs are defined, that is, if variants are counted as seen or unseen. The work by Maldonado et al. (2017) defines seen MWEs – similarly to our understanding – as MWEs present in the test set and sharing the same dependency structure, POS tags and lemmas with at least one annotated MWE in the training set. They find out that most systems perform better when the proportion of seen MWEs is high. This result suggests that MWE variant identification remains a hard problem and deserves being addressed explicitly by dedicated methods.

3 VMWE variation in French

VMWEs are known to have specific lexical and morpho-syntactic *variability profiles*: they are more or less variable, but usually not as variable as regular phrases with the same syntactic structure. Therefore, methods used for detecting paraphrases of regular phrases (Fujita and Isabelle, 2015) cannot be straightforwardly applied to VMWEs. Different aspects of variability may be considered.

Firstly, MWE-hood is, by nature, a lexical phenomenon, that is, a particular idiomatic reading is available only in presence of a combination of particular lexical units. Replacing one of them by a semantically close lexeme usually leads to the loss of idiomatic reading, e.g. (6) is an idiom but (7) can

only be understood literally. Lexical variability is admitted by some VMWEs, but usually with a very restricted list of equivalents only, as in (11) and (12). In this work, unlike Fazly et al. (2009), we exclude lexical variability from our scope. More precisely, two VMWE occurrences are considered *morpho-syntactic variants* of each other if they have the same (idiomatic) meaning and are *lexically identical*, that is, the multisets of the lemmas of their lexicalized components are identical. For instance, (12) is considered a separate VMWE rather than a variant of (11). Also, (16) is not a variant of (13) due to the determiner which is lexicalized in the latter but not in the former. Note that textually identical occurrences of a VMWE, like (13) vs. (15) are also considered variants.

(6) *Léa* ***tourne la page*** '*Léa turns the page*'⇒'Léa stops dealing with what occupied her'
(7) *Léa pivote la page* '*Léa rotates the page*'; *Léa tourne la feuille* '*Léa turns the sheet*'
(8) ***tournera****-t-elle* ***la page*** *de la politique?* '*turn-will-she the page of the politics?*'⇒'Will she stop dealing with politics?'
(9) ***la page*** *a été* ***tournée*** '*the page has been turned*'⇒'the subject is no longer dealt with'
(10) *elle tourne les pages de la politique* '*she turns the pages of politics*'
(11) *Ses plaintes me* ***cassent les oreilles*** '*his moans break my ears*'⇒'his moans annoy me'
(12) *Ses plaintes me* ***cassent les pieds*** '*his moans break my feet*'⇒'his moans annoy me'
(13) ***faire le tour*** *du sujet* '*make the tour of the topic*'⇒'consider all aspects of the topic'
(14) ***refaire le tour*** *du sujet* '*remake the tour of the topic*'⇒'reconsider all aspects of the topic'
(15) *on va en* ***faire le tour*** '*we will make the tour of it*'⇒'we will consider all aspects of it'
(16) ***faire*** *un* ***tour*** *en famille* '*make a tour in family*'⇒'go out with one's family'
(17) *le tour fait 3 kilomètres* '*the tour makes 3 kilometers*'⇒'the tour is 3 km long'
(18) ***filer le*** *parfait/grand.ADJ* ***amour*** '*spin the perfect/great love*'⇒'to live a perfect/great love'
(19) *Léa* ***prenait*** *souvent* ***la porte*** '*Léa took often the door*'⇒'Léa was often forced to go out'
(20) *la porte de la maison a été prise par Léa* '*the door of the house was taken by Léa*'

Another aspect of VMWE variability is morphological: some VMWE components do not admit inflection, as the noun *page* '*page*' in (8) vs. (10), while others do, as the verb *tourner* '*turn*' in (8). In rare cases, the head noun accepts derivation, often with prefixes like *re-*, to express repetition, as in (14).

Syntactic variability is also crucial for the VMWE variability profile. Firstly, idiomatic readings most often correspond to literal readings used metaphorically. Therefore, the VMWE components have to occur in a syntactic configuration which relates to the literal reading. For instance, (17) is not a variant of (13), since *tour* takes a different semantic role. Still, both the literal and the idiomatic meaning can sometimes be preserved under syntactic variation as in (6) vs. (9). Secondly, some types of syntactic variation (e.g. passivization) tend to be exhibited less frequently by VMWEs than by non-VMWEs of the same syntactic structures (Fazly et al., 2009). Thirdly, some types of syntactic dependencies can be specific to some VMWEs, e.g. (18) involves a compulsory though non-lexicalized adjectival modifier of the noun. This also means that this VMWE admits insertions of external elements between its lexicalized components. Finally, as previously shown (Pasquer et al., 2018), syntactic features are particularly strong indicators for linguistically motivated similarity and flexibility measures of (French) VWMEs.

The main challenge in modeling the variability profiles of VMWEs lies in the fact that VMWEs of the same syntactic structure may behave differently. For instance, the VMWEs in (6) and (19) both admit interrogation, insertions and inflection of the verb, and prohibit inflection of the noun. Still, modification of the noun and passivization is exhibited by the former (8)–(9) but not by the latter (20).

4 Corpus

We study VMWE variability on a corpus of French texts annotated with VMWEs for the PARSEME shared task.[3] In addition to VMWE annotation (based on universal guidelines), the corpus contains POS, lemmas, morphological features and dependency structures, which we use in our experiments. The original release of the VMWE-annotated corpus contained two sub-corpora: the French part of the Universal Dependencies v1.4 corpus (Nivre et al., 2016) and Sequoia (Candito and Seddah, 2012).

[3] `http://multiword.sf.net/sharedtask2017`, corpus at `http://hdl.handle.net/11372/LRT-2282`

Corpus	All POS patterns				All Verb-(Det-)Noun		Verb-(Det-)Noun variants	
	# Sentences	# Tokens	# VMWEs	# occ.	# VMWEs	# occ.	# VMWEs	# occ.
TrC	17,880	450,221	1,584	4,462	854	2098	n/a	n/a
TeC	1,667	35,784	291	500	177	283	86	132

Table 1: Number of different VMWE types (# VMWEs) and their token occurrences (# occ.) per POS and variability class in the training corpus (TrC) and in the test corpus (TeC).

The tagsets and dependency trees used in both sub-corpora were incompatible. We homogenized them by replacing the released annotations with their UDv2-compatible versions, while VMWE annotations remain unchanged.

The VMWEs annotated for French belong to three main categories: inherently reflexive verbs (***se plaindre*** '*pity oneself*'⇒'complain'), light-verb constructions (***prendre** une **décision*** '*take a decision*') and idioms (***tourner la page*** '*turn the page*'⇒'stop dealing with what one was occupied with'). In this study we focus only on verb-noun combinations, possibly involving a lexicalized determiner, hence only the last two VMWE categories are relevant.

Tab. 1 shows the VMWE statistics of the training corpus (TrC) and test corpus (TeC). Note that, with our focus on variants of seen VMWEs having the Verb-(Det-)Noun structure, only 86 VMWEs and their 132 occurrences from TeC are to be predicted. Among them, only 35 occurrences (26.5%) appear under identical surface forms in TeC as in TrC.

5 Baseline variant identification

The aim of our study is to automatically identify morpho-syntactic variants of VMWEs in a syntactically parsed French corpus. More precisely, given a set of VMWEs annotated in TrC, we wish to identify all their morpho-syntactic variants in TeC. Each of such variants to be predicted in TeC will be called a seen-in-train variant (STV). Note that, a given expression e in TeC can be an STV only if it has the same multiset of lemmas as a VMWE e' from TrC. But this condition is not sufficient due to the existence of literal readings and accidental co-occurrences (Sec. 1). We hypothesize that distinguishing such spurious candidates from STVs should be possible by relying on the morpho-syntactic variability profile. In other words, e has good chances to be a morpho-syntactic variant of e' if both are morpho-syntactically similar.

This idea suggests a relatively straightforward baseline STV identification approach, similar to the *BagOfDeps* strategy proposed by Savary and Cordeiro (2018). Given a dependency-parsed TeC, we extract each set of nodes e so that: (i) the multiset of lemmas in e is identical to some VMWE e' in TrC, (ii) e forms a connected dependency graph. Note that this approach neutralizes both morphological variation (via lemmatization) and syntactic variation (we disregard the labels and directions of the dependencies). As a result, the predicted STVs can be either true positives (Fig.1.a and 1.b) or false positives (Fig.1.d). False negatives include occurrences without direct connection like complex determiners (Fig.1.f). The baseline is also sensitive to syntactic annotation errors, as in Fig.1.e (*rôle* '*role*' rather than *clé* '*key*' should be the head of *rôle clé* 'key role'). Despite its simplicity, the baseline is already very strong, as shown in the first line of Tab. 2. It extracts 158 Verb-(Det-)Noun candidates, which – compared to the 132 STVs to be extracted – yield the F-score of 0.88. These results notably confirm previous observations that literal readings are rare (Waszczuk et al., 2016; Savary and Cordeiro, 2018).

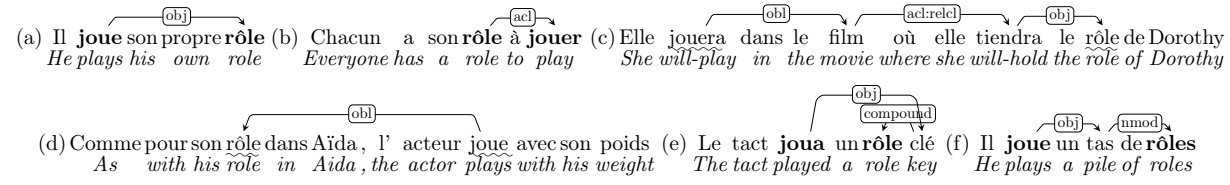

Figure 1: True positive (a,b) and true negative (c) STVs. False positive STV (d), inconsistent syntactic annotation (e), indirect dependencies (e,f)

6 Variant identification as a classification problem

We hypothesize that VMWE variability profiles can be exploited to achieve a better VMWE variant identification than the baseline. Variability profiles will be approximated by a set of morpho-syntactic features (Sec. 6.1). We will first use a more exhaustive candidate extraction method than the baseline, so as to avoid false negatives. Then, the features should help discriminate true STVs from spurious candidates (Sec. 6.2).

6.1 Variation-related features

We describe the profile of any given STV candidate expression e (hereafter called 'candidate' or e) by binary features characterized according to two dimensions:
Morphological, linear and syntactic features model characteristics discussed in Sec. 3, such as verbal/nominal inflection, linear insertions between components (e.g. adjectives, adverbs) or discrepancies between syntactic dependencies (e.g. subject vs. object).
Comparative and absolute features: Comparative features are based on the comparison of e with lexically identical VMWEs from TrC, e.g. if e's noun is found at least once with same inflection in a lexically identical VMWE in TrC, then the COMP_genderNumber(e) feature is True. Absolute features consist in contextual properties of e, e.g. if e's noun is in singular, the numberSing(e) feature is True.

6.1.1 Comparative features (COMP)

COMP features verify the correspondence of e's properties with those of VMWEs annotated in TrC. Henceforth, we say that a property of e matches if it is the same for e and for at least one lexically identical annotated VMWE in TrC. COMP features are binary (their value is False when no match is found and True otherwise) and belong to 3 types: morphological, linear and syntactic.
Morphological similarity is represented by a single feature, COMP_genderNumber, which focuses on the nominal inflection only, since previous experiments showed that it is more informative than verbal inflection (Pasquer et al., 2018). Thus, COMP_genderNumber(e) is True if e's inflection matches.
Linear similarity features account for different degrees of similarity:

- COMP_insertRaw: True if the POS sequence of insertions between lexicalized components of e matches. For instance, if e has an Adj-Adj-Det insertion, it can only match a VMWE having the identical sequence of insertions, Adj-Adj-Det.
- COMP_insertWithoutDuplicate: Same as above, ignoring duplicated POS, so Adj-Adj-Det can match both Adj-Adj-Det and Adj-Det. This feature neutralizes duplications, as we believe that MWEs have constraints on the POS of allowed insertions, but not on their number (e.g. it is rare that one adjective is tolerated but not two).
- COMP_insertPartial: True if a POS substring of more than 50% and less than 100% of insertions in e matches. This feature identifies the similarity between ***play*** *a*.DET *very*.ADV *important*.ADJ ***role*** and ***play*** *his*.DET *first*.ADJ *major* ***role***. However, linear similarity does not capture postnominal adjectives (e.g. ***jouer*** *un* ***rôle*** *important* 'play a role important'), hence the interest in syntactic similarity.

Syntactic similarity Like for morphological features, we noticed greater influence of the outgoing dependencies of the noun, as in Fig.1.b (contrary to incoming dependencies in Fig.1.a), hence the focus on the noun. These syntactic features include:

- COMP_depSynNounTotal, COMP_depSynNounPartial: True if 100%, or more than 50% and less than 100%, of the noun's outgoing dependencies match, respectively.
- COMP_distSyn : True if the syntactic distance between the verb and the noun matches. The *syntactic distance* between two components is defined as the number of elements in the syntactic dependency chain between these two components, regardless of the direction of the dependencies and excluding the components themselves. Inside known VMWEs, the syntactic distance never exceeds 2, so that the value is set to ">2" when this case occurs. In Fig. 1.c this chain is composed of *jouera-film-tiendra-rôle*. Supposing that all the VMWEs of the corpus are represented in Fig. 1, the syntactic distance is 2, hence COMP_distSyn = False since it differs from all the annotated VMWEs.

- COMP_typeDistSyn: True if the type of the syntactic distance matches. This type takes dependency directions into account, and can be serial (in Fig. 1.f, the noun *rôles* '*roles*' depends on *tas* '*pile*', which itself depends on the verb *jouer* '*play*'), parallel (in Fig. 1.e, both *jouer* '*play*' and *rôle* '*role*' depend on the non-lexicalized component *clé* '*key*') or "nonEvaluated" when the syntactic distance exceeds 2 elements.

We expect COMP features to be highly relevant. For instance, the noun in many VMWEs must remain in singular, as in (8) vs. (10), or in plural, as in (11). However, COMP features might fail in case of rare VMWEs, and cannot be calculated for hapaxes. Therefore, absolute features should also be useful.

6.1.2 Absolute features (ABS)

For each candidate, the set of absolute features includes the following.

Lemmas of the components can be considered individually or as a sequence. When considered individually, the VMWE in (6) has ABS_verbLemma = *tourner*, ABS_nounLemma = *page*, and ABS_detLemma = *le* (when there is no lexicalized determiner, ABS_detLemma = noDet). When considered as a sequence, we sort lemmas lexicographically to neutralize variable word order, as in (6) vs. (9), and obtain the *Normalized Form* (NF).[4] Thus, *le;page;tourner* 'the;page;turn' is the ABS_NF of the VMWEs in (6)–(10).

Morphological features for the noun are: ABS_numberSing, ABS_numberPlur, ABS_genderMasc, ABS_genderFem. Their value is True when the respective property is satisfied. For the verb, we only consider derivational inflection. Namely, the ABS_verbPrefix feature is True only if the verb starts with one of the most frequent repetition (*re/ré/r*) or negation prefixes (*de/dé*) and if the verb without this prefix matches the verb in any VMWE annotated in TrC, e.g. ABS_verbPrefix is True for the VMWE in (14).

Linear features correspond to insertions between lexicalized components. We add one insertion feature per possible POS, which is set to True if the e has an insertion with the given POS. For instance in ***effectuer*** *une*.DET *bonne*.ADJ *première*.ADJ ***saison*** 'make a good first season', we obtain ABS_insert_DET = True, ABS_insert_ADJ = True, ABS_insert_ADV = False, etc.

Syntactic features are defined similarly to the linear ones. We introduce one feature per possible dependency relation, and set it to True if the noun in e has at least one outgoing dependency arc with this relation (e.g. ABS_relDepNoun_amod = True for the previous example).

Syntactic distance between the verb and the noun is defined as before, i.e. as the number of elements in the syntactic dependency chain except for the components themselves.

Type of syntactic distance also follows the one in COMP features, and can be either serial or parallel. COMP features are far less numerous than ABS features (9 vs. 72 in the example detailed in App. A).

6.2 Overview of the method

Our goal is to extract STVs, i.e. variants of previously seen VMWEs. We wish to enhance over the baseline (Sec. 5), which is sensitive to indirect dependencies, to dependencies irrelevant to variation, and to annotation errors (which may cause a dramatic performance drop with automatically parsed data), as shown in Fig. 1.d–f. We thus propose an enhanced approach in two steps, named *mweVIDE* for MWE variant identification. First, a more relaxed variant candidate extraction is designed so as to achieve a better recall than the baseline (usually at the expense of a dramatic drop in precision), by extracting, for each annotated e' in TrC, its lexically identical candidates, disregarding syntax. Then, the extracted candidates are filtered by a binary classifier based on the features described above. Fig.2 illustrates the two most crucial phases of this process: the candidate extraction and the feature extraction. VMWE candidates from TrC are numbered (TrC1)–(TrC9), as shown in the first gray box. Those from TeC, (TeC1)–(TeC5), are shown in the first white box. Training of the classifier itself, and its application to prediction are standard and disregarded in Fig.2.

6.2.1 Variant candidate extraction

The easiest way to obtain a larger variant candidate extraction is to search, in each sentence, the simultaneous presence of all components of known VMWEs whatever their order, as described by Savary

[4]The drawback of NF is that it ignores nominal inflection that could help distinguish between (very rare) VMWEs with different surface forms (e.g. ***fermer l'oeil*** '*close the eye*'⇒'sleep' vs. ***fermer les yeux*** '*close the eyes*'⇒'turn a blind eye').

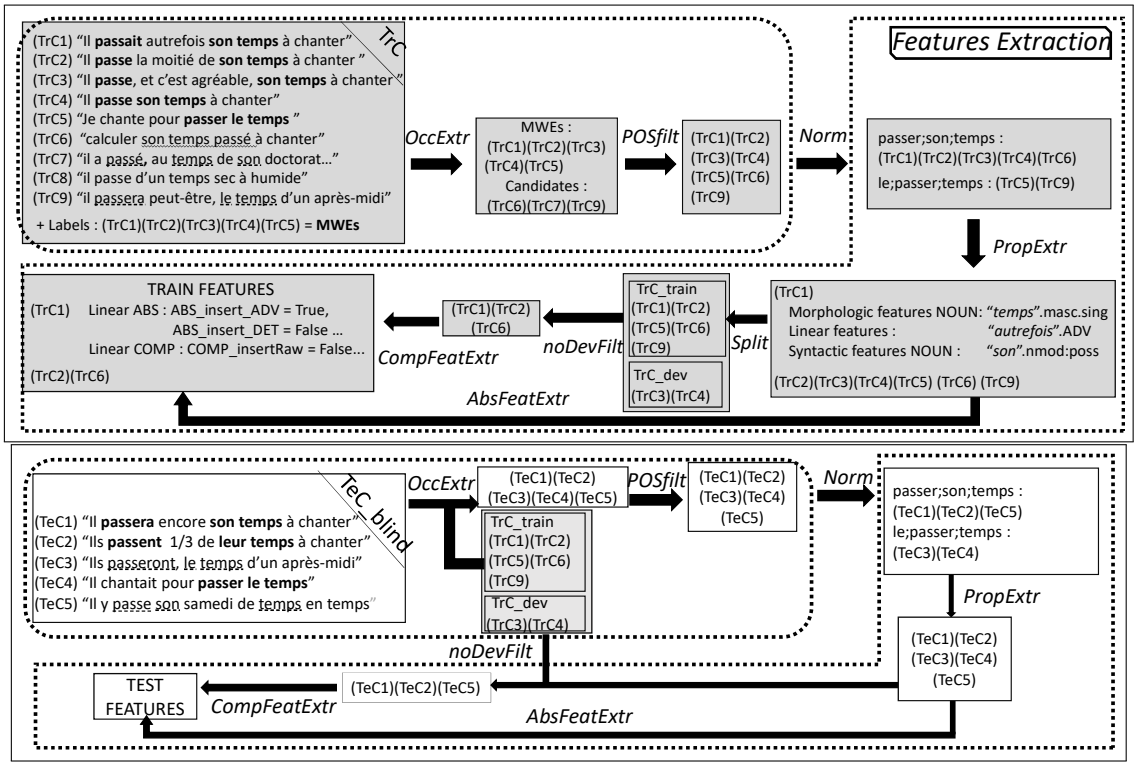

Figure 2: Candidate and feature extraction in training (above) and during prediction (below).

and Cordeiro (2018). This phase is called *OccExtr*. Consider the first gray box in Fig.2, representing TrC, and several expressions related to two VMWEs ***passer son temps*** 'spend one's time' and ***passer le temps*** '*spend the time*'⇒'invent activities when one has nothing to do'. (TrC1)–(TrC5) are true VMWEs, (TrC6) is a literal reading, (TrC7) and (TrC9) are accidental co-occurrences, and (TrC8) is not lexically identical to any of the two VMWEs due to a missing determiner. Since *OccExtr* matches only occurrences lexically identical to known VMWE, it will extract, trivially, (TrC1)–(TrC5), and (TrC6)(TrC7)(TrC9), but not (TrC8).[5] *OccExtr* is expected to improve recall but precision should also significantly decrease. To limit possibly huge levels of noise, *OccExtr* extracts candidates in the shortest window. For instance, in (TeC5) only one candidate is extracted despite the repetition of the noun *temps* '*time*'.[6] *OccExtr* is enhanced by additional filtering (*POSfilt*) which eliminates irrelevant POS orders (i.e. incorrect in French), such as Verb-Noun-Det in (TrC7). Finally, the NFs for all candidates are produced (*Norm*).

6.2.2 *mweVIDE*: Binary classification of candidates

As a result of the previous phase, we obtain sets of candidates grouped by their NFs (cf. upper-right-hand box). Morpho-syntactic properties for these candidates are then extracted (*PropExtr*) and the whole set is randomly divided (*Split*) into two equal subsets TrC_train and TrC_dev, so as to be able to calculate COMP values. For a given candidate e from TrC_train, when no VMWE with the same NF is annotated in TrC_dev, e is deleted (*noDevFilt*), as is the case of (TrC6) and (TrC9). This may lead to a loss of almost 30% of TrC_train but is inevitable, since for each candidate we need at least one lexically identical VMWE in TrC_dev to calculate the values of the comparative features. The final set of positive and negative candidates is then represented by COMP and ABS features (*CompFeatExtr* and *AbsFeatExtr*), and fed into classifier training.

In the prediction phase (lower part of Fig. 2), the candidates are extracted from TeC on the basis of their

[5]Note that *OccExtr* in training (upper part of Fig.2) applies to TrC itself, since classification calls for both positive and negative examples. Conversely, in prediction, *OccExtr* is applied to the blind version of TeC.

[6]This may lead to both false negatives and false positives, as in *la* ***collaboration*** *que nous avons.[AUX=avoir 'have']* ***eue****.[VERB=avoir 'have']* 'the collaboration that we have had'.

lexical identity (i.e. NF-identity) with annotated VMWEs from TrC. They are *POSfilt*ered, *Norm*alized and subject to *Prop*erty *Extr*action, as in training. But no *Split* needs to be performed, since each TeC candidate can be compared to a VMWE in TrC_dev. The 3 candidates remaining after *noDevFilt* are then represented as COMP and ABS features and classified into STVs or non-STVs.

We use the NLTK Naive Bayes classifier.[7] In the training phase it takes as input the set of positive and negative examples represented as feature-value pairs, together with their classes (STV/non-STV), and outputs a model. In the prediction phase, it takes as input the trained model and the candidates to classify, represented as feature-value pairs, and for each candidate outputs an STV or a non-STV label. Training and prediction are repeated 10 times with random TrC_train/TrC_dev splits, corresponding to half of TrC each. The classification performance (recall, precision, F1-measure) is evaluated on the manually annotated version of TeC. For each of the 10 training and prediction turns, we also obtain the 100 most informative features associated to the (non-)STV label prediction.

7 Results

We firstly compare the performance of *OccExtr* with the baseline. Table 2 shows the performances of the systems on TrC (2098 occurrences) and TeC corpora (132 occurrences). As expected, the precision of

	TrC #candidates	R	P	F1	TeC #candidates	R	P	F1
Baseline	2414	0.970	0.85	0.91	158	0.97	0.81	0.88
OccExtr	5364	0.999	0.39	0.56	484	1.00	0.27	0.43
mweVIDE	5364	n/a	n/a	n/a	484	0.97 (σ_R = 0.007)	0.87 (σ_P = 0.015)	0.92 (σ_{F1} = 0.010)

Table 2: Performance measure of the baseline, OccExtr and mweVIDE on TrC and TeC (average on 10 random TrC_train vs. TrC_dev splits)

OccExtr is significantly lower than that of the baseline. Conversely, both recall scores remain very close at a high level of performance. An almost perfect recall is reached by *OccExtr*, but its very low F-measure (0.56 and 0.43 on TrC and TeC, respectively) prevents its direct application to variant detection.

The classification of the candidates implemented in mweVIDE aims at improving the precision of *OccExtr* without impacting recall. As shown in Table 2, mweVIDE increases the F1-measure to 0.92 (vs. 0.88 for the baseline) with a better precision (0.87 vs. 0.81). Recall value is similar to the baseline. The low standard deviation of the F-score (0.01) we observed on the 10 experiments is a good indication of the statistical significance of this improvement. Note that, for this evaluation, we do not have access to data annotated directly with ground truth, i.e. describing which VMWEs are variants of each other. We rely instead on the comparison between the predicted and annotated VMWEs, hence a possible bias because of (i) distinct VMWEs with identical NFs, (ii) similar VMWEs with distinct NFs (due lemmatization errors).

8 Error analysis

In this section we perform a qualitative analysis of the errors performed by the baseline and mweVIDE, in order to investigate what kinds of situation are correctly handled by each system, and which leave room for improvement. The baseline cannot identify components without direct dependency (Fig.1.f) or with parallel dependencies (Fig.1.e). Other omitted cases include: (i) literal readings, (ii) overlap with another VMWE, (iii) coordination, (iv) enumeration, (v) co-reference : ***réunion*** *qui interviendra après celles* ***tenues*** 'meeting which will occur after those held' (where *those* = *meeting*), even though this specific case should not have been annotated according to the annotation guidelines.

As to mweVIDE, it may suffer from the absence of similar VMWEs in TrC (e.g. neither similar syntactic distance, nor insertions or nominal inflection). After 10 evaluations on TeC, the predicted (non-)STV labels are constant in 98.9% of the cases. Moreover, despite the *noDevFilt*, 99.2% of the candidates

[7] NLTK (`http://www.nltk.org/`) is used with default parameter values, including ELEProbDist (Expected Likelihood Estimation based on Laplacian smoothing with alpha value = 0.5). Preliminary experiments with a linear SVM classifier (NLTK SklearnClassifier) yielded slightly lower performances than Naive Bayes. Given the low amount of the training data, Naive Bayes seems a satisfactory trade-off compared to other classification methods which require more data (e.g. polynomial SVM).

STV label feature	$\overline{ratio}$	σ_{ratio}	non-STV label feature	$\overline{ratio}$	σ_{ratio}
COMP_insertWithoutDuplicate=True	28.7	5.3	ABS_insert_VERB=1	123.7	53.5
COMP_insertRaw=True	28.4	5.8	COMP_typeDistSyn=False	121.5	38.7
ABS_NF=jouer;rôle	14.8	3.6	ABS_typeDistSyn=parallel	71.4	39.5
ABS_nounLemma=fin	10.7	4.8	ABS_insert_CCONJ=1	70.4	30.9
ABS_nounLemma=face	9.9	2.5	COMP_distSyn=False	61.0	23.4
ABS_NF=fin;mettre	8.7	2.2	ABS_distSyn=2	56.4	20.1
ABS_distSyn=0	8.7	0.8	ABS_insert_SCONJ=1	52.5	18.3
COMP_distSyn=True	8.6	0.8	ABS_insert_PUNCT=1	29.5	7.9
ABS_NF=faire;partie	8.0	1.6	ABS_distSyn=1	27.9	8.4
ABS_verbLemma=jouer	7.6	2.5	ABS_insert_PROPN=1	25.3	8.2

Table 3: The most informative features (COMP in gray) according to the averaged likelihood ratio for (non-)STV prediction over 10 tests.

could be classified at least once. Among the 19 false positives, 5 should actually have been annotated as VMWEs. Cases of coordination, never identified by the baseline, are partially identified by mweVIDE, which seems to better handle the omission of a coordination conjunction than its addition. The better precision is due to our POS filtering which excludes irrelevant patterns and mainly to the typology of insertions in known VMWEs used by the classifier. For instance, subordination between the components is less frequent in VMWEs than in coincidental co-occurrences, as will be exposed in the next section.

9 Most informative features

Likelihood ratios associated with each feature allow us to determine which features are the most discriminative in distinguishing STV vs. non-STV candidates. For instance, if the ratio is equal to 2 for a given feature and label, this means that this feature is twice more frequently associated with this label than with the complementary label in the training set. By averaging these ratios on the 10 evaluations, we can extract the ten most informative ones for prediction (Tab. 3). The ranking presented in Tab. 3 is given with mean and standard deviation ratio values: the standard deviation is a good indication that, whatever the selected TrC_train corpus, the same features tend to favor STV (resp. non-STV) labels.

9.1 Features relevant for the STV label

As shown in Tab. 3, features which favor the STV label are both of COMP and ABS type.

COMP features The insertion-related features (COMP_insertRaw=True and COMP_insertWithoutDuplicate=True) are always ranked first. This is understandable, since with fewer insertions the number of potential inserted POS sequences is lower. In TrC, 97% of the VMWEs have less than 5 inserted elements vs. 30% of false STV candidates, which reflects how much non-STVs are susceptible to be more variable in terms of insertions. An identical syntactic distance (COMP_distSyn=True) often means that its value is 0 in a seen VMWE and its STV. False STVs tend to overpass this value because the syntactic connection between elements is often looser than inside a VMWE.

ABS features The relevance of ABS_distSyn=0 (which is the only criterion of the baseline) can be explained in the same way as COMP_distSyn=True above. The other most informative features are NFs (like *jouer;rôle* '*play role*') or isolated components (only the verb *jouer* '*play*'), which means that they are less frequently associated with coincidental co-occurrences or literal readings.

9.2 Features relevant for the non-STV label

Contrary to the prediction of the STV label, COMP features have negative values for non-STV labels, which highlights how much similarity (respectively dissimilarity) with known VMWEs is an important criterion for the STV (resp. non-STV) prediction.

COMP features The only discriminative COMP features are the syntactic distance and its type, when different from known VMWEs.

ABS features Insertions and syntactic distance are the most discriminating. All elements that tend to break the link between components are relevant. In descending order, it can be the insertion of a verb, a

coordinating or subordinating conjunction, a punctuation or a proper noun. As for syntactic features, the most relevant here are the syntactic distance higher than zero or parallel syntactic distance.

10 Conclusions and future work

We developed a system for Verb-(Det-)Noun VMWE variant identification based on morphological and syntactic profiles of candidates which permit their classification as (non-)variant thanks to a set of absolute and comparative features (the latter relying on the comparison with known VMWEs). The system shows satisfactory performance (F1-measure = 0.92) and good reproducibility. This represents an improvement over the baseline based on the presence of a syntactic connection between components. Morphology does not appear as a discriminating feature to label a candidate as a STV, contrary to comparative linear features (e.g. similar POS insertions), which were rarely taken into account in previous works (Fazly et al., 2009). In general, comparative features appear to be more relevant for STV prediction. Conversely, absolute features predominate in non-STV variant prediction, since non-STVs often exhibit unreferenced variation profiles. Given the reduced size of our test corpus, we need further evaluations on larger corpora to improve our variant identification system, notably to handle cases of coordinations that are not systematically identified. We also project to include other comparative features such as orthographic similarity (***show one's true colours/colors***) or lexical similarity (***take a bath/shower***), which might help classify rare VMWEs that could not be evaluated here. A wider range of patterns than only Verb-(Det-)Noun should also be considered. Finally, future extension of our system to other languages available in the PARSEME corpus, whether Romance or not, could also be considered.

Acknowledgements

This work was funded by the French PARSEME-FR grant (ANR-14-CERA-0001). We are grateful to the anonymous reviewers for their useful comments.

References

Isabelle Augenstein, Leon Derczynski, and Kalina Bontcheva. 2017. Generalisation in named entity recognition. *Comput. Speech Lang.*, 44(C):61–83, July.

Timothy Baldwin and Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkhya and Fred J. Damerau, editors, *Handbook of Natural Language Processing*, pages 267–292. CRC Press, Taylor and Francis Group, Boca Raton, FL, USA, 2 edition.

Marie Candito and Djamé Seddah. 2012. Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. In *Proceedings of TALN 2012*, juin.

Matthieu Constant and Joakim Nivre. 2016. A transition-based system for joint lexical and syntactic analysis. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2016)*, pages 161–171, Berlin.

Matthieu Constant, Joseph Le Roux, and Anthony Sigogne. 2013. Combining compound recognition and PCFG-LA parsing with word lattices and conditional random fields. *TSLP Special Issue on MWEs: from theory to practice and use, part 2 (TSLP)*, 10(3).

Mathieu Constant, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner, and Amalia Todirascu. 2017. Multiword expression processing: A survey. *Computational Linguistics*, 43(4):837–892.

Afsaneh Fazly, Paul Cook, and Suzanne Stevenson. 2009. Unsupervised type and token identification of idiomatic expressions. *Computational Linguistics*, 35(1):61–103.

Atsushi Fujita and Pierre Isabelle. 2015. Expanding paraphrase lexicons by exploiting lexical variants. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 630–640. Association for Computational Linguistics.

Spence Green, Marie-Catherine de Marneffe, and Christopher D. Manning. 2013. Parsing models for identifying multiword expressions. *Computational Linguistics*, 39(1):195–227.

Gaston Gross. 1988. Degré de figement des noms composés. *Langages*, 90:57–72.

Ben Hachey, Will Radford, Joel Nothman, Matthew Honnibal, and James R. Curran. 2013. Evaluating Entity Linking with Wikipedia. *Artif. Intell.*, 194:130–150, January.

Christian Jacquemin, 2001. *Spotting and Discovering Terms through Natural Language Processing*. The MIT Press.

Cvetana Krstev, Ivan Obradovic, Milos Utvic, and Dusko Vitas. 2014. A system for named entity recognition based on local grammars. *J. Log. Comput.*, 24(2):473–489.

Alfredo Maldonado, Lifeng Han, Erwan Moreau, Ashjan Alsulaimani, Koel Dutta Chowdhury, Carl Vogel, and Qun Liu. 2017. Detection of verbal multi-word expressions via conditional random fields with syntactic dependency features and semantic re-ranking. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 114–120, Valencia, Spain, April. Association for Computational Linguistics.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association (ELRA), may.

Geoffrey Nunberg, Ivan A. Sag, and Thomas Wasow. 1994. Idioms. *Language*, 70:491–538.

Caroline Pasquer, Agata Savary, Jean-Yves Antoine, and Carlos Ramisch. 2018. Towards a Variability Measure for Multiword Expressions. In *NAACL*, New Orleans, United States, June.

Ivan Sag, Timothy Baldwin, Francis Bond, Ann Copestak, and Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, page 1–15, Mexico City,Mexico.

Agata Savary and Silvio Ricardo Cordeiro. 2018. Literal readings of multiword expressions: as scarce as hen's teeth. In *Proceedings of the 16th Workshop on Treebanks and Linguistic Theories (TLT 16)*, Prague, Czech Republic.

Agata Savary and Christian Jacquemin. 2003. Reducing Information Variation in Text. *LNCS*, 2705:145–181.

Agata Savary, Carlos Ramisch, Silvio Cordeiro, Federico Sangati, Veronika Vincze, Behrang QasemiZadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova, and Antoine Doucet. 2017. The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 31–47, Valencia, Spain, April. Association for Computational Linguistics.

Nathan Schneider, Emily Danchik, Chris Dyer, and Noah A. Smith. 2014. Discriminative lexical semantic segmentation with gaps: running the MWE gamut. *Transactions of the ACL*, 2:193–206.

Nathan Schneider, Dirk Hovy, Anders Johannsen, and Marine Carpuat. 2016. Semeval-2016 task 10: Detecting minimal semantic units and their meanings (dimsum). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 546–559. Association for Computational Linguistics.

Livnat Herzig Sheinfux, Tali Arad Greshler, Nurit Melnik, and Shuly Wintner, 2017. *Verbal MWEs: Idiomaticity and flexibility*, pages 5–38. Language Science Press, à paraître.

Agnès Tutin. 2016. Comparing morphological and syntactic variations of support verb constructions and verbal full phrasemes in French: a corpus based study. In *PARSEME COST Action. Relieving the pain in the neck in natural language processing: 7th final general meeting*, Dubrovnik, Croatia.

Veronika Vincze, János Zsibrita, and István Nagy. 2013. Dependency parsing for identifying Hungarian light verb constructions. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 207–215.

Jakub Waszczuk, Agata Savary, and Yannick Parmentier. 2016. Promoting multiword expressions in A* TAG parsing. In *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 429–439.

A Absolute and comparative features for a sample STV candidate.

STV candidate in TeC:

- *Elle **fit**.VERB alors.ADV la.DET **connaissance**.NOUN de son futur mari.* '*She made then the knowledge of her future husband*'⇒'She got then acquainted with her future husband'

4 annotated VMWEs in TrC (basis for COMP features):

- *Le film [...] permet de **faire**.VERB la.DET **connaissance**.NOUN d'une kyrielle de personnages.* '*The movie allows to make the knowledge of a myriad of characters*'⇒'The movie allows one to get acquainted with a myriad of characters'
- *Ricky **fait**.VERB la.DET **connaissance**.NOUN de Junito* '*Ricky make the knowledge of Junito*'⇒'Ricky gets acquainted with Junito'.
- *On pense qu'il y **fit**.VERB la.DET **connaissance**.NOUN de Jean Heynlin et de Guillaume Fichet* 'One thinks that he made the knowledge of Jean Heynlin and Guillaume Fichet'It is believed that he got acquainted with Jean Heynlin and Guillaume Fichet
- *Lorsque Jacques exprima le souhait de **faire**.VERB la.DET **connaissance**.NOUN de les autres leaders de le groupe* 'When Jacques expressed the desire to make teh knowledge of the other leaders of the group'When Jack expressed the desire to get acquainted with the other leaders of the group

The table illustrates the calculation of the ABS and COMP features for these examples. ABS_insert_ is followed by the part-of-speech of the considered insertions, and ABS_relDepNoun_ by the outgoing dependencies related to the noun. ∅ means lacking information whereas "_" means that it is underspecified.

COMP features	**ABS features**	**ABS features**
'COMP_depSynNounPartial': True	'ABS_NF': 'connaissance;faire'	'ABS_insert_ADV': True
'COMP_depSynNounTotal': False	'ABS_detLemma': 'noDET'	'ABS_insert_AUX': False
'COMP_genderNumber': True	'ABS_nounLemma': 'connaissance'	'ABS_insert_CCONJ': False
'COMP_insertWithoutDuplicate': False	'ABS_verbLemma': 'faire'	'ABS_insert_DET': True
'COMP_insertRaw': False	'ABS_verbPrefix': 'noPrefix'	'ABS_insert_INTJ': False
'COMP_insertPartial': True	'ABS_insert_ _': False	'ABS_insert_NUM': False
'COMP_distSyn' : True	'ABS_insert_ADJ': False	'ABS_insert_PART': False
'COMP_typeDistSyn': True	'ABS_insert_ADP': False	'ABS_insert_PRON': False
	'ABS_insert_PROPN': False	'ABS_insert_PUNCT': False
	'ABS_insert_SCONJ': False	'ABS_insert_SYM': False
	'ABS_insert_VERB': False	'ABS_insert_X': False
	'ABS_genderFem': True	'ABS_genderMasc': False
	'ABS_numberPlur': False	'ABS_numberSing': True
	'ABS_relDepNoun_acl:relcl': False	'ABS_relDepNoun_acl': False
	'ABS_relDepNoun_advcl': False	'ABS_relDepNoun_advmod': False
	'ABS_relDepNoun_amod': False	'ABS_relDepNoun_appos': False
	'ABS_relDepNoun_aux:pass': False	'ABS_relDepNoun_aux': False
	'ABS_relDepNoun_case': False	'ABS_relDepNoun_cc': False
	'ABS_relDepNoun_ccomp': False	'ABS_relDepNoun_compound': False
	'ABS_relDepNoun_conj': False	'ABS_relDepNoun_cop': False
	'ABS_relDepNoun_csubj': False	'ABS_relDepNoun_dep': False
	'ABS_relDepNoun_det': True	'ABS_relDepNoun_discourse': False
	'ABS_relDepNoun_dislocated': False	'ABS_relDepNoun_expl': False
	'ABS_relDepNoun_fixed': False	'ABS_relDepNoun_flat:foreign': False
	'ABS_relDepNoun_flat:name': False	'ABS_relDepNoun_goeswith': False
	'ABS_relDepNoun_iobj': False	'ABS_relDepNoun_mark': False
	'ABS_relDepNoun_nmod:poss': False	'ABS_relDepNoun_nmod': False
	'ABS_relDepNoun_nsubj:pass': False	'ABS_relDepNoun_nsubj': False
	'ABS_relDepNoun_nummod': False	'ABS_relDepNoun_obj': False
	'ABS_relDepNoun_obl': False	'ABS_relDepNoun_orphan': False
	'ABS_relDepNoun_parataxis': False	'ABS_relDepNoun_punct': False
	'ABS_relDepNoun_reparandum': False	'ABS_relDepNoun_root': False
	'ABS_relDepNoun_vocative': False	'ABS_relDepNoun_xcomp': False
	'ABS_distSyn': False	'ABS_typeDistSyn': 'serial'