A Language-Independent Approach to Extracting Derivational Relations from an Inflectional Lexicon

Marion Baranes, Benoît Sagot

A Language-Independent Approach to Extracting Derivational Relations from an Inflectional Lexicon

Marion Baranes[1,2] **Benoît Sagot**[2]

1. viavoo, 77 rue de Paris, 92100 Boulogne-Billancourt, France
2. Alpage, INRIA & Université Paris Diderot, bâtiment Olympe de Gouges, 75013 Paris, France
marion.baranes@viavoo.com, benoit.sagot@inria.fr

Abstract

In this paper, we describe and evaluate an unsupervised method for acquiring pairs of lexical entries belonging to the same morphological family, i.e., derivationally related words, starting from a purely inflectional lexicon. Our approach relies on transformation rules that relate lexical entries with the one another, and which are automatically extracted from the inflected lexicon based on surface form analogies and on part-of-speech information. It is generic enough to be applied to any language with a mainly concatenative derivational morphology. Results were obtained and evaluated on English, French, German and Spanish. Precision results are satisfying, and our French results favorably compare with another resource, although its construction relied on manually developed lexicographic information whereas our approach only requires an inflectional lexicon.

Keywords: Formal Analogy, Morphological Analysis, Derivational Relation

1. Introduction

Derivational morphology can provide useful information for many natural language processing tasks. Indeed, it can improve any application which has to deal with unknown words in general and neologisms in particular, such as information extraction, spell-checking and others. For example, Bernhard et al. (2011) have shown that it can improve question answering systems. Sagot et al. (2013) use derivational and compound analysis for morphological lexicon extension, in order to determine if an unknown word should be used to create a new lexical entry. Derivational morphology can also help extending lexical resources with syntactic (e.g., sub-categorisation frames) and semantic information (e.g., wordnets). One could for example infer sub-categorisation frames for nouns which are derived from known verbs for which sub-categorisation information is known.

We define a *morphological family* as a set of semantically related lexical entries which differ by their prefix and/or suffix, thus limiting ourselves to concatenative derivational morphology. For example, the English words *learn*, *unlearn* or *learner* share a common lexical basis, and just differ with respect to derivational affixes. Therefore, they belong to the same morphological family. We shall denote as *derivationally related* two morphological lexical entries that belong to the same morphological family.

Our system performs an analogy-based unsupervized extraction of weighted transformation rules that relate derivationally related lexical entries, and use these rules for extracting derivational relations within an existing inflectional lexicon.[1] Our transformation rules can also be used to infer morphological information (both inflectional and derivational) for wordforms unknown to the inflectional lexicon. Our system is language-independent, although restricted to concatenative derivational morphology. We have evaluated it on four languages, namely English, French, German and Spanish.

After a brief overview of the related work in Section 2, we describe our system in Section 3. In Section 4, we introduce our experimental setup and the resulting derivational resources. Finally, we describe and discuss quantitative evaluation results (Section 5).

2. Related Work

Even though morphological analysis is the subject of numerous studies, it does not give rise to many lexical resources which contain derivational relations between lexical entries. Among the four languages at hand, although we can cite CELEX (Burnage, 1990) for English, German and Dutch, French seems to have received the most attention in that regard. Some resources try to join French verbs to their related nominals. For example, *VerbAction* (Tanguy and Hathout, 2002) pairs verbs with their derivationally related action nouns (*accuser* 'accuse' – *accusation* 'accusation'), and *VerbAgent* (Tribout et al., 2012) with derivationally related agent nouns (*accuser* 'accuse' – *accusateur* 'accuser'). Others studies try to do the opposite and pair denominal nouns with their derivationally related verbs, such as Nomage (Balvet et al., 2011) for French or NOMLEX (Macleod et al., 1998) for English. For French as well, DenALex (Strnadová and Sagot, 2011) pairs denominal adjectives with their nominal basis (*atome* 'atom' – *atomique* 'atomic'), whereas the database MORDAN (Koehl, 2013) contains 3983 pairs of deadjectival nouns and their base adjective. We can also cite POLYMOTS (Gala et al., 2010), a lexical resource that groups wordforms in morphological families and provides some information about their internal structure. However, the development of all these resources was based on manual work and/or on manually built lexical information (dictionary definitions, derivational

[1] We define an inflectional lexicon as a set of entries of the form (citation form, part-of-speech, inflection class) together with a morphological grammar, which allows for generating (citation form, inflected form, part-of-speech, morphological tag) tuples.

morphological grammars, etc). Other studies propose less supervised systems for detecting morphological families or, more generally, for acquiring morphological information. Can and Manandhar (2009) describe an usupervised approach based that relies, among others, on parts-of-speech in order to produce a morphological analysis. Bernhard (2010) describes two distinct unsupervised systems (*MorphoClust* and *MorphoNet*). The first one is based on hierarchical clustering methods, the second one uses a graph-based algorithm in order to group related wordforms of the same family in a lexical network. We can also cite Goldsmith (2001), Kazakov and Manandhar (2001), Neuvel and Fulop (2002), Creutz and Lagus (2005), Monson et al. (2007) or, in a more general context, works which participate to the Morpho Challenge[2]. In this paper, however, we focus on learning by analogy which is an approach used for morphological systems (Lepage, 2000; Stroppa and Yvon, 2005; Lavallée and Langlais, 2009). Hathout (2010) introduces Morphonette, a system restricted to French which uses morphological similarity, lexicographic definition-based semantic similarity and analogy. This system obtains results which are quite close to ours for French. This is why we choose to use it as a reference for our evaluation.

3. Our approach

Our computational system for derivational analysis is inspired by works based on formal analogy. Analogy aims at relating two pairs of terms: x is to y as z is to t, written $x : y :: z : t$. In order to match lexical entries which belong to the same morphological family, we look for affixation rules which are shared by both entries. Our system only requires an inflectional lexicon in order to function properly, i.e., an inventory of lexical entries, each associated with its list of inflected forms. For example, one can infer that the English adjective *liable* belongs to the same morphological family as *liability* if we find that: (1) *liability* is known, e.g., it is in the lexicon and (2) there is a rule which allows us to substitute the suffix *–ability* with *–able*.

This analysis requires the acquisition of such rules, which we call *transformation rules*.[3]

As we aim at relating lexical entries from an inflectional lexicon to the one another, we could try and extract these transformation rules directly on their citation forms. However, it is often the case that the inflected forms of a same lexical entry are based on more than one stem (e.g., in French, *aller* 'go', *vais* '(I) go/am going', *irai* '(I) will go'). Derivational processes do not necessarily use as their starting point the stem that underlies the citation form (e.g., the French noun *ouverture* 'opening' is related to the verb *ouvrir* 'open', but can be considered as based on the same stem as its past participle *ouvert* 'open'). Therefore, we first extract transformation rules which relate inflected forms with citation forms within the lexicon, we then infer relations between inflected forms and lexical entries (through their citation forms), and finally replace in these relations inflected forms by their lexical entry, thus building a set relations between lexical entries.

Let us first describe how we extract our transformation rules. We achieve this using a 4-step algorithm:

1. we extract a preliminary set of generic rules than can only be either purely prefixal or purely suffixal;
2. we generate a first set of (inflected form, citation form) pairs based on these generic rules;
3. we extract from these pairs a new set of rules, which can be prefixal and/or suffixal rules and that include POS information;
4. we extract (inflected form, citation form) pairs based on this final set of rules, and then merge these (inflected form, citation form) pairs into pairs relating lexical entries.

We shall now describe these 4 steps in more detail.

3.1. Extracting the preliminary set of rules

The aim of this step is to learn prefixes and suffixes particular to the language concerned. In other words, we extract preliminary transformation rules that are either prefixal or suffixal, not both. To do that, we proceed in two stages:

1. The extraction of prefixal rules: in order to extract these rules, we pair all inflected forms with all citation forms of our lexicon. For each possible pair of the form (inflected form, citation form), we compare their structure. If these two forms differ only by their prefixes and share a significant common part (at least 3 characters),[4] we create a rule that relates the prefix of the inflected form to that of the citation form. These rules contain the input and output prefix as well as a short context (the first common letter, which immediately follows both prefixes). For example, given the English (inflected form, citation form) pair *subtitle-title*, the extracted rule will be $\{sub \rightarrow \_\}\{t\}$ (i.e., before a *t*, replace *sub* by the empty string, represented by the symbol '_').
2. The extraction of suffixal rules: this extraction is the mirror image of the extraction of prefixal rules. We extract all (inflected form, citation form) pairs that differ only by their suffixes and share a common part. Whatever the language processed, this common part have to be at least 6 characters[5]. Reducing this threshold increases significantly the run time (the

[2] http://research.ics.aalto.fi/events/morphochallenge/

[3] These transformation rules need not be derivation rules. They only model transformations that relate two words belonging to the same morphological family, be them directly derivationaly related or not, and independently of the possible direction of such a derivational relation.

[4] This threshold was set emprically because of running time issues when using lower values.

[5] Afterwards, in order to choose this threshold according to a language, a non-supervised method will have to be implement. Indeed, we realised that the number of prefixal rules, which has been extracted with this threshold, ranges from less than 100,000 (with the English) to more than 6 millions (with the French).

threshold 5 is three times slowlier than the threshold 6) and, at same time, the many additional prefixal rules, which could be obtained in this way, are virtually never correct. Given such a pair, we extract a suffixal rule in the same way as we do for prefixal rules. For example, given the English (inflected form, citation form) pair *laughs-laughing*, the extracted rule will be $\{h\}\{s \rightarrow ing\}$.

In the process, we count the number of times each such rule is extracted. Each rule is stored together with its number of occurrences.

3.2. Generating the preliminary set of (inflected form, citation form) pairs

We use the preliminary transformation rules extracted as described in the previous section for relating inflected forms with citation forms. To do that, for each inflected form in our lexicon we try and apply all applicable prefixal rules (i.e., rules such that the inflected form starts with the input prefix followed the one-character context specified in the rule). We also try and apply to the input inflected form all applicable suffixal rules. We finally try and apply to the input inflected form all pairs of applicable rules, one prefixal and one suffixal. Each time the result is a citation form known to the lexicon, we store the pair (original inflected form, resulting citation form). These pairs constitute a preliminary derivational lexicon, which relates inflected forms to citation forms, including cases where both the prefix and the suffix are changed during the transformation. For example, if we have the Spanish inflected form *abono* 'subscription', we will join it, among other things, to the citation form *desabono* 'cancellation of subscription' with the prefixal rule {_→des}, to the citation form *abonar* 'to subscribe' with the suffixal rule {o→ar} and to the citation form *desabonar* 'cancel a subscription' with this two latters affixal rules.

Two such pairs, say (x, y) and (z, t), are in an analogical relation, $x : y :: z : t$ if they have been obtained using the same transformation rule(s). At this stage, our set of derivational relations is still noisy as shown in the table 1.

PAIRS	PREFIX	SUFFIX
appreciable → depreciate	{ap → de}{p}	{i}{able → ate}
appreciable → precis	{ap → _}{p}	{able → s}
appreciable → appreciably	_	{e → y}
appreciablest → appreciable	_	{est → e}
demoded → modest	{de → _}	{ed → est}
demoded → modish	{de → _}	{ed → ish}

Table 1: Example of preliminary pairs extracted from our English lexicon

3.3. Transformation rules extraction

Based on the preliminary derivational lexicon, generated as described in the previous section, we extract a new set of transformation rules, which can now be simultaneously prefixal and suffixal. We extract from each pair (x, y) in our preliminary derivational lexicon the prefixal and/or suffixal transformation rules which transforms x into y.

We also consider as part of the rule several additional pieces of information:

- First, depending on whether the citation form of the inflected form x is the same as (the citation form) y or not, we mark the rule as inflectional or as derivational. Rules extracted from the English pairs *mediatisations-mediatisation* or *upgraded-upgrade*, for example, will be labelled as inflectional, wheareas rules extracted from English pairs such as *accused-unaccusable* or *communication-communicate* will be labelled as derivational.

- Second, as our lexicon provides us with the part-of-speech and the morphological features (e.g., gender, number) for the inflected form x, as well as the part-of-speech and inflection class for the citation form y, we incorporate this information within our rules.

- Third, we store for each rule the number of distinct (inflected form, citation form) pairs in the preliminary derivational lexicon from which it was extracted; this figure will be considered as the number of occurrences of the rule.

As a result, these rules can be writen as follows:

$$\textit{(prefix, suffix, POS, morph. feat.)} \xrightarrow{\textit{infl./der.}} \textit{(prefix', suffix', POS', infl. class.')}$$

In order to minimize the noise in the extracted rules, we discard all rules which appear less than 80 times — an empirically-chosen value that we shall discuss in Section 5.1.. A few examples from our French, English, German and Spanish data are shown in table 2.

3.4. Derivational relations extraction

Once this final set of rules is extracted, we can relate morphological entries from our inflectional lexicon with the one another. As in section 3.2. above, but using the final set of transformation rules, we relate pairs of the form (x, y), where x is an inflected form (with its part-of-speech and morphological tag) and y is a lexical entry (citation form, part-of-speech, inflection class), provided the rule transforms x into y, while respecting part-of-speech and morphological information at both ends. Next, we replace x by its lexical entry, thus creating pairs of lexical entries (hopefully) belonging to the same morphological family. Table 2 contains a few examples of pairs extracted with our transformation rules.

Note that a morphological lexicon does not distinguish between the different senses of polysemous words that behave in the exact same manner at the morphological level. Therefore, it might be the case that we relate two morphological entries eventhough this relationship only applies to specific senses of both morphological entries involved. This is because morphological families are not defined only in morphological terms, but also involve semantic affinities.

LANGUAGE	CATEGORY_MORPHTAG	PREFIX	SUFFIX	OCC	EXAMPLE
German	adj_plain.pl.nom.primary.long → n_sg.gen.short	_	ische→ie	136	morphologische → morphologie
German	v_subj.pres.sg.1.long → v_inf.long	_ →be	_	105	denke → bedenken
English	A_ → R_inf	_→ un	_→ly	1123	fortunate → unfortunately
English	N_ → V_inf	_	tion→te	1123	evacuation → evacuate
French	adj_Kfp → v_W	_	ées → er	6483	données → donner
French	nc_ms → nc_ms	_	ement → age	828	chiffrement → chiffrage
Spanish	v_MN0000 → n_CMS000	a→_	ar→o	342	abalear → baleo
Spanish	n_CMS000 → v_MN0000	_	o→ar	1665	trabajo → trabajar

Table 2: A few transformation rules extracted from various inflectional lexicons

4. Experimental setup and results

We have implemented the language-independent algorithm described in the previous section and have applied it to inflectional lexicons for English, German, Spanish and French. More precisely, we have used the large-scale inflectional lexicons developed for these languages within the Alexina framework (Sagot, 2010), namely EnLex for English, DeLex for German (Sagot, 2014), the L*effe* for Spanish (Molinero et al., 2009) and the L*efff* for French (Sagot, 2010). We only retained nominal, verbal, adjectival and adverbial entries. The size of the resultig lexicons, more precisely their number of inflected forms and lexical entries, are shown in the Table 3.

LEXICON	LANGUAGE	INFLECTED FORMS	LEXICAL ENTRIES
EnLex	English	463,576	181,494
DeLex	German	398,096	58,841
L*effe*	Spanish	694,040	101,417
L*efff*	French	446,432	59,617

Table 3: Number of the entries in the input inflectional lexicons

Table 4 shows the number of rules and pairs obtained for each language. It provides the number of transformation rules we extracted as explained in Section 3.3., the number of pairs of related lexical entries we created with these rules, and the ratio of the number of created pairs with respect to the number of all possible pairs in each lexicon (i.e., the number of lexical entries times this number minus 1).

LANGUAGE	TRANF. RULES	PAIRS OF LEXICAL ENTRIES	EXTRACTED PAIRS / POSSIBLE PAIRS
English	11,748	597,148	0.015 ‰
German	6,812	10,639	0.017 ‰
Spanish	6,000	69,694	0.005 ‰
French	8,834	84,927	0.003 ‰

Table 4: Number of transformation rules relating an inflected for to a lexical entries, together with the number of pairs of lexical entries extracted for each language based on these rules. See text for details.

We define the part-of-speech pattern of a relation between two lexical entries as the pair consisting of the part-of-speech of its input entry and the one of its output entry. Table 5 illustrates, for each language, the most common part-of-speech patterns in the derivational lexicons we have acquired. Percentages in this table indicate the share of relations that follow the corresponding part-of-speech pattern. For instance, 18% of the English derivational pairs we have acquired relate two nouns with one another. The non-homogeneous distribution of our part-of-speech patterns for the four languages at hand can be explained at least in part by the fact that each lexicon displays a different distribution of parts-of-speech accross lexical entries — this is at least in part a property of the respective languages involved, but is also certainly influenced by the properties of each lexicons. For example, in English, the most frequent parts-of-speech are nouns and adjectives. It is therefore no surprise that part-if-speech patterns involving these categories cover as much as 66% of all patterns. French and Spanish lexicons display similar distributions of parts-of-speech accross lexical entry, which results in the fact that their part-of-speech patterns are similarly frequent. As far as German is concerned, the massive amount of *v → adj* patterns is a consequence of the fact that our German lexicon, DeLex, (temporarily) models present participle as adjectival lexical entries[6] and not, as one could expect, as inflected form of the corresponding verbal lemma. As a result, our algorithm creates a large amount of pairs relating verbal lexical entries with the adjectival entries that covers their present participle forms.

5. Evaluation

We performed two kinds of evaluation. First, we evaluated our results on all 4 languages at hand by manually assessing the quality of the extracted pairs of lexical entries. Second, we evaluated our French results againt the morphological resource Morphonette (Hathout, 2010), in order to assess the recall of our system, the overlap between both resources and the precision of pairs found respectively in our data only, in Morphonette only or in both resources.

5.1. Manual evaluation of the precision

In order to evaluate the precision of our pairs of lexical entries, we extracted for each language 100 randomly selected such pairs. We have manually associated each of them with one of the following tags:

CORR: Both lexical entries belong to the same morphological family (e.g., English *preconfiguring – configured*);

[6] DeLex is still under development. Future versions will integrate present participle forms as part of the verbal paradigms.

POS PATTERN	ENGLISH	GERMAN	SPANISH	FRENCH	EXAMPLE
adj → adj	17%	4%	0.4%	1%	*adaptable → adaptative*
adj → n	16%	11%	13%	6%	*random → randomization*
n → adj	15%	3%	10%	8%	*compression → compressible*
n → n	18%	1%	32%	32%	*self-destruction → self-destructive*
n → v	5%	2%	10%	17%	*abandonment → to abandon*
v → adj	7%	52%	6%	5%	*to sanction → unsanctioned*
v → n	8%	21%	17%	20%	*to tabulate → tabulation*
v → v	4%	7%	5%	4%	*labeled → mislabel*

Table 5: Most frequent POS-based patterns for each language.

UNUSUAL: The derivational relation between both lexical entries is correct, but only applies to senses that are rare for at least one of the lexical entries at hand (e.g., French *tentement* '(fencing) striking the adversary's sword twice with one's own' – *tenter* '(fencing) perform a *tentement*', whereas the most common sense is 'try');

DIACHR: Both lexical entries share indeed a common etymology, but cannot be synchronically considered as belonging to the same family (e.g., French *mariner* 'stew' – *marin* 'sailor').

INFLEX: Both lexical entries belong clearly to the same morphological family but they are linked by an inflexional relation and not a derivational relation (e.g., English *congratulation* – *congratulations*);

ERR: The two lexical entries do not belong to the same morphological family (e.g., French *graver* 'engrave' – *grave* 'serious');

LANG.	CORR	UNUSUAL	DIACHR	INFLEX	ERR
English	98	1	0	0	1
German	98	0	1	0	1
Spanish	73	8	2	4	13
French	89	2	3	4	2

Table 6: Evaluation results for pairs of lexical entries

The results of this evaluation are shown in Table 6. One can notice that error rates for English, French and German are rather low (between 1% and 2% — although such figures are to be taken with care given the low amount of pairs evaluated), whereas the error rate for Spanish is higher. This is caused by a noisy rule which creates derivationnal links between words which are distinguished by the suffixes {ear#} and {ar#}. This rule creates wrong derivationnal pairs such as *zapar* 'sap, mine' – *zapear* 'shoo (a cat)' or *copear* 'drink (familiar)' – *copar* 'corner (the market)', and is the cause of 9 our the 13 pairs tagged as ERR.[7].

Results of this evaluation espacially depend on the quality of the rules used to create our pairs of lexical entries. As mentioned in Section 3.3., we chose to retain only rules with a frequency at least 80. This choice of this threshold is the result of a two-step analysis, which we shall now sketch.

First, we performed an qualitative analysis of our rules according to their frequency. Our goal here is to determine from which threshold our rules become reliable. To do that, we have randomly extracted 50 rules at each frequency level between 10 and 80 that is a multiple of 10. In other words, we have evaluated 50 rules with a frequency of 10, 50 rules with a frequency of 20, and so on up to 80. For each such set of rules, we evaluated whether at least one pair of lexical entries belonging to the same morphological familly could be built and would be related by the rule at hand. Results for French are given in Figure 1. It shows that a sufficient level of precision can not be expected with a threshold of approximatively 50 or below.

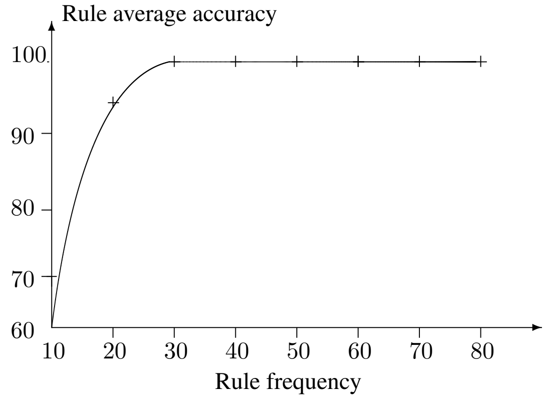

Figure 1: Evaluation of transformation rules according to their occurrence number

Second, as a result of this first step, we evaluated our system on French taking into account all rules with a frequency of 50 or higher (rather than 80 or higher). This allowed us to obtain a new set of 156,064 derivational pairs (vs. 84,927 with a threshold of 80). We have evaluated a set of 100 randomly selected derivational pairs. Out of these 100 pairs, we tagged 74 as CORR (correct), 1 as DIACHR, 15 as INFLEX and 10 tagged as ERR. In other words, 26 out of 100 pairs are incorrect in some way. The set of underlying transformation rules is therefore much less reliable than with a threshold of 80.

[7] Our Spanish result could be easily improved by deleting this noisy rule.

5.2. Comparative evaluation of our French results with another resource

Next, we compared our French lexical entry pairs to the resource Morphonette (Hathout, 2010) which was also built based on analogy, but leveraging previous manual lexicographic work. For comparison purposes, we ignore inflectional class information of our results, as they are not available in Morphonette, and only retain the part-of-speech of each entry. Moreover, we remove from Morphonette all inflectional relations and all relations involving neoclassical compounds (e.g., *psychopathe* – 'psychopath'), as our approach only targets formally regular derivational phenomena. This results in 76,754 pairs (out of 96,081 initially). This being done, both datasets have 22,591 pairs in common, 62,336 pairs are created only by our system, whereas 54,163 pairs are only found in Morphonette. To compare our respective datasets, we randomly extracted and manually evaluated 200 pairs of related lexical entries for each of these three subsets. Results are presented in Table 7.

SYSTEM	CORRECT	OTHER
Pairs found both in Morphonette and in our system	97.5%	2.5%
Morphonette only	96.5%	3.5%
Our system only	94%	6%

Table 7: Accuracy of our system with respect to Morphonette

Our system and Morphonette share about one third of their results in common, which is not much: The recall of both approaches is still relatively low. If we consider the union of both datasets as the reference, which is obviously optimisic, one can compute recall figures for both datasets: we would then reach a recall of 61%, whereas Morphonette would have a recall of 55%.

Our precision rates are almost as high as those of Morphonette, which is very satisfactory for at least two reasons. First, our system has produced more pairs of lexical entries. Second, it is important to recall that Morphonette does rely on a massive amount of manually built lexicographic information, as it takes advantage of the electronic version of the large-scale dictionary *Trésor de la Langue Française*, by exploiting the lexicographic definitions it contains. Our system manages to reach accuracy levels which are almost as high as Morphonette's without exploiting any such costly information. As a result, our system is language-independent — provided derivational morphology can be considered concatenative —, and we applied it indeed to four different languages, whereas developing an version of Morphonette for another language would require the use of a large-scale electronic dictionary for that language.

6. Conclusion

In this paper, we introduced an unsupervised language-independent system which automatically extracts transformation rules and derivationally related pairs of lexical entries from an inflectional lexicon, without the need for any manual supervision. We tested and evaluated it for precision on four languages, namely English, French, German and Spanish. In addition, we performed a comparative evaluation of our French results with the morphological resource Morphonette. Despite the fact that the construction of Morphonette was based on the exploitation of manually built lexicographic definitions, we reach almost the same level of precision and a slightly higher recall with our purely unsupervized approach. In the future, we would like to integrate our results for providing our inflectional lexicons with a derivational layer, e.g., in the form of information about the morphological family of each entry.

Acknowledgements

This work was partly funded by the French ANR national grant EDyLex (ANR-09-CORD-008).

7. References

Balvet, Antonio, Barque, Lucie, Condette, Marie-Hélène, Haas, Pauline, Huyghe, Richard, Marín, Rafael, and Merlo, Aurélie. (2011). Nomage: an electronic lexicon of French deverbal nouns based on a semantically annotated corpus. In *Proceedings of the First International Workshop on Lexical Recources (WoLeR 2011)*, pages 8–15, Ljubljana, Slovenia.

Bernhard, Delphine, Cartoni, Bruno, and Tribout, Delphine. (2011). A Task-based Evaluation of French Morphological Resources and Tools. *Linguistic Issues in Language Technology*, 5(2).

Bernhard, Delphine. (2010). Apprentissage non supervisé de familles morphologiques: Comparaison de méthodes et aspects multilingues. *Traitement Automatique des Langues*, 51(2):11–39.

Burnage, Gavin. (1990). Celex: A guide for users. Technical report, University of Nijmegen, Center for Lexical Information.

Can, Burcu and Manandhar, Suresh. (2009). Unsupervised learning of morphology by using syntactic categories. In *Proceedings of 10th Workshop of the Cross-Language Evaluation Forum, CLEF'2009*, Corfu, Greece.

Creutz, Mathias and Lagus, Krista. (2005). Inducing the morphological lexicon of a natural language from unannotated text. In *Proceedings of the International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning (AKRR'05)*, pages 106–113.

Gala, Nuria, Rey, Véronique, and Zock, Michael. (2010). A tool for linking stems and conceptual fragments to enhance word access. In *Proceedings of the Seventh international conference on Language Resources and Evaluation (LREC'2010)*, Valletta, Malta.

Goldsmith, John. (2001). Unsupervised learning of the morphology of a natural language. *Computational Linguistics*, 27(2):153–198.

Hathout, Nabil. (2010). Morphonette: a morphological network of French. *CoRR*, abs/1005.3902.

Kazakov, Dimitar and Manandhar, Suresh. (2001). Unsupervised learning of word segmentation rules with genetic algorithms and inductive logic programming. *Machine Learning*, 43(1-2):121–162.

Koehl, Aurore. (2013). Une base de données des noms désadjectivaux du français : le modèle mordan. In *Proceedings of Corpus et Outils en Linguistique, langues et parole*, Strasbourg, France.

Lavallée, Jean François and Langlais, Philippe. (2009). Unsupervised morphological analysis by formal analogy. In *Proceedings of Workshop of the Cross-Language Evaluation Forum (CLEF 2009), Lecture Notes in Computer Science*, pages 618–625, Corfu, Greece.

Lepage, Yves. (2000). Languages of analogical strings. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, pages 488–494, Saarbrücken, Germany.

Macleod, Catherine, Grishman, Ralph, Meyers, Adam, Barrett, Leslie, and Reeves, Ruth. (1998). Nomlex: A lexicon of nominalizations. In *Proceedings of the 8th EURALEX International Congress*, pages 488–494, Liege, Belgium.

Molinero, Miguel A., Sagot, Benoît, and Nicolas, Lionel. (2009). A morphological and syntactic wide-coverage lexicon for Spanish: The Leffe. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2009)*, Borovets, Bulgarie.

Monson, Christian, Carbonell, Jaime, Lavie, Alon, and Levin, Lori. (2007). Paramor: Minimally supervised induction of paradigm structure and morphological analysis. In *Proceedings of the Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology (SIGMORPHON 2007).*

Neuvel, Sylvain and Fulop, Sean A. (2002). Unsupervised learning of morphology without morphemes. *CoRR*, cs.CL/0205072.

Sagot, Benoît, Nouvel, Damien, Mouilleron, Virginie, and Baranes, Marion. (2013). Extension dynamique de lexiques morphologiques pour le français à partir d'un flux textuel. In *Proceedings of TALN'2013*, pages 407–420, Les sables d'Olonne, France.

Sagot, Benoît. (2010). The Le*fff* , a freely available and large-coverage morphological and syntactic lexicon for French. In *Proceedings of the Seventh international conference on Language Resources and Evaluation (LREC 2010)*, Valletta, Malta.

Sagot, Benoît. (2014). DeLex, a freely-avaible, large-scale and linguistically grounded morphological lexicon for German. In *Proceedings of the 9th international conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland.

Strnadová, Jana and Sagot, Benoît. (2011). Construction d'un lexique des adjectifs dénominaux. In *Proceedings of TALN'2011*, pages 69–74, Montpellier, France.

Stroppa, Nicolas and Yvon, François. (2005). An analogical learner for morphological analysis. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL'2005*, pages 120–127, Stroudsburg, PA, USA.

Tanguy, Ludovic and Hathout, Nabil. (2002). Webaffix: un outil d'acquisition morphologique dérivationnelle à partir du Web. In *Proceedings of TALN'2002*, pages 245–254, Nancy, France.

Tribout, Delphine, Ligozat, Anne-Laure, and Bernhard, Delphine. (2012). Constitution automatique d'une ressource morphologique : VerbAgent. In *Proceedings of CMLF'2012*, Lyon, France.