Evaluating Morphological Resources: a Task-Based Study for French Question Answering

Delphine Bernhard, Bruno Cartoni, Delphine Tribout

Evaluating Morphological Resources: a Task-Based Study for French Question Answering

Delphine Bernhard[1], **Bruno Cartoni**[2], **Delphine Tribout**[1]

(1) LIMSI-CNRS, 91403 Orsay, France
(2) Département de linguistique, Université de Genève, Suisse
bernhard@limsi.fr, bruno.cartoni@unige.ch, tribout@limsi.fr

Abstract

Morphology is a key component for many Natural Language Processing applications. In this article, we focus on one prototypical application, namely Question Answering (QA). In QA, morphological relations, especially those relying on the derivation and compounding processes, are often addressed in a superficial manner. Considering that some resources are able to provide deep and precise knowledge about a large spectrum of morphological processes, the issue lies first in determining the morphological phenomena which are most relevant for QA systems and second in evaluating the coverage of existing resources in this respect. To this aim, we describe a manual annotation and analysis of French question-answer pairs, which was performed in order to produce a unique and well-characterised reference dataset. Based on this study, we evaluate five different morphological resources for French and show that some dedicated resources are still lacking, which would cover phenomena such as denominal adjectives and agent deverbal nouns.

1. Introduction

Morphological resources are central to many Natural Language Processing applications. Despite their importance, resources are still lacking for many languages and domains, in particular with regard to constructional morphology, i.e. derivation and compounding. Moreover, they are usually evaluated intrinsically by human evaluators. As for extrinsic evaluations, they focus on the performance gains which can be obtained by using morphological knowledge in a specific applications, e.g. speech recognition (Creutz et al., 2007), machine translation (Koehn and Hoang, 2007) or information retrieval (Hahn et al., 2003). In this article, we propose a new method for evaluating resources which consists in manually building a task specific gold-standard in order to measure the coverage and quality of morphological resources. Here we focus on one prototypical application, namely Question Answering (QA).

QA systems aim at providing a precise answer to a given user question. To this aim, they usually rely on an Information Retrieval (IR) component which attempts to match words in the question and words in the text passages containing a potential answer. The major difficulty lies in the lexical gap problem, which occurs when a document is related to a question even though it does not contain the same words as the question. QA and IR systems must thus find a way of retrieving relevant documents without relying only on mere identity between words. In this context, morphology has often been preferred over semantics because the integration of morphological knowledge is easier. Research in IR and QA has thus tried to incorporate morphological knowledge, either by expanding the query, by indexing documents with morphologically motivated units or by using question reformulation or rephrasing patterns to identify the answer.

Most of the research carried out so far made use of simple heuristic-based stemming techniques which cut off word endings (such as (Lennon et al., 1988), (Harman, 1991), (Fuller and Zobel, 1998)). These turned out to be rather efficient for languages with a "less-rich" morphology, such as English, but they are not available for all languages (McNamee et al., 2009). In most cases, the recall is slightly improved, but these techniques also produce some noise, as shown by the example described in Bilotti et al. (2004): *organisation* and *organ* are stemmed to the same form by the Porter Algorithm. Another interesting piece of research, described in Moreau and Claveau (2006), shows that extending the query by morphological knowledge significantly improves the results, in most of the European languages for which they performed the experiment. To acquire morphological knowledge, they made use of a learning method based on analogy techniques. Consequently, they captured only affixation processes, and moreover only transparent affixation processes (that share a rather long character string), leaving aside conversion, reduction processes, or affixation on suppletive forms. They also admitted that, even with some precautions (long minimal character string, etc.), some incorrect pairs of morphologically related words are captured (*pondre* with *répondre*).

As we have shown, QA applications mostly rely on partial or superficial morphological knowledge. Moreover, only few studies specifically address the role of morphology within such systems. Most of the evaluations are extrinsic (based on the measurement of the improvement of an entire system when a morphological "module" is applied), and globally, the use of morphology (either indexing or query expansion) is very coarse.

However, some morphological resources are now able to provide detailed and precise knowledge about a large spectrum of morphological processes. The issue is more in weighting the relations to be implemented, and in determining the resources to be used – or built if lacking. Hence, we address two specific research questions in this article:

1. What morphological phenomena are most relevant in a QA application?
2. How well do available resources for French morphol-

ogy cover these phenomena?

These two aspects are linked together because we first need to characterize the morphological relations wich are relevant in a QA task in order to evaluate the use of existing morphological resources in a QA system.
We therefore performed our evaluation of morphological resources for French in two steps. First, we have manually annotated and analysed pairs made of a question on one side, and the snippet containing the answer on the other side, in order to determine the morphological relations involved. Secondly, we used this set of pairs of morphologically related words as a gold-standard to evaluate the coverage of available resources for French. Since the gold-standard has been carefully characterised, precise measures can be computed for different morphological processes.
The contributions of the paper are as follows:

- We present the constitution and the analysis of a unique gold-standard for morphological relations, based on a detailed annotation of three different corpora of question-answer pairs. This study provides important insights on the type of morphological knowledge to be integrated into QA systems in order to improve their performance.
- We evaluate and compare five different morphological resources for French, including both inflectional and derivational morphology.
- We show that resources covering some important morphological phenomena are still lacking for the French language and make concrete proposals about the resources which would be most helpful for QA.

2. Annotation of Question-Answer pairs

2.1. Description of the datasets

The datasets gathered for the annotation come from three very different QA corpora: Quæro, EQueR-Medical and Conique, which are presented below. Our aim in annotating different types of corpora was to determine if there are significant differences in the morphological processes observed depending on the type of data. Table 1 presents statistical information on each corpus.

Quæro The French Quæro corpus has been built for QA evaluation (Quintard et al., 2010) whithin the Quæro project. The corpus consists of 2.5M French documents extracted from the web and a set of 250 questions for the 2008 evaluation and 507 questions for the 2009 evaluation. The document corpus has been constituted by taking the first 100 pages returned by the Exalead search-engine for a set of requests found in the search-engine's logs. As for the questions, they have been written by French native speakers by using the contents of the documents for the 2008 evaluation, and by using only the query logs of the search-engine for the 2009 evaluation. There are three types of questions: factual questions, boolean questions which ask for a yes-no answer and questions requiring a list for answer.
We have constituted our corpus for the annotation task by taking all the snippets returned by the Ritel-QA System (Quintard et al., 2010) that have been manually validated as containing the correct answer for each factual question of the two evaluation campaigns. We thus obtained 566 pairs of question and snippet containing the answer, 338 from the the 2008 evaluation and 228 from the 2009 evaluation.

EQueR-Medical The EQueR evaluation dataset has been constituted within the EQueR-EVALDA evaluation campaign for French Question Answering systems (Ayache et al., 2006). The campaign included two main tasks: (i) general domain QA over a collection of newspaper articles and senate reports and (ii) specialised domain QA over a collection of medical texts. We restricted our annotation study to the medical questions. The answer snippets were retrieved by the participant systems and manually validated by a specialised judge.
Overall, the EQueR-Medical dataset comprises 394 question answer-snippet pairs for 200 different questions.

CONIQUE The CONIQUE corpus has been built with the objective of studying relevant answer justifications for QA systems (Grappy et al., 2010). Answer justifications provide additional material to the user, so that she/he may trust the answer retrieved by the system. The corpus is based on a subset of 291 questions from the French EQueR campaign (Ayache et al., 2006) and several CLEF campaigns. Candidate answer snippets have been retrieved from the French Wikipedia using a coarse retrieval mechanism and manually annotated by seven annotators. In contrast to the two previously described datasets, answer snippets in CONIQUE do not correspond to the output provided by QA systems. It therefore constitutes an almost full recall dataset, devoid of any bias inherent to QA systems such as high question-snippet token overlap.
We automatically pre-processed the annotated corpus to retrieve question-snippet pairs. We only kept full or partial justifications. Moreover, we reduced the snippet to up to three sentences, centred on an annotated justification. Overall, the dataset we annotated comprises 664 question-answer pairs, for 201 different questions.

2.2. Annotation methodology

The annotation was manually performed by three trained independent annotators,[1] using the YAWAT alignment tool (Germann, 2008). YAWAT was originally developed to align words in bilingual sentence-pairs for machine translation evaluation. In our case, we aligned words and phrases in question-answer pairs and typed their morphological relation. We defined three tags for morphological relations: one for inflection, another for derivation and another for compounding. Since there can be more than one morphological step between two morphologically related words we defined specific guidelines for the annotation.
First, we did not annotate inflectional variants of auxiliaries and determiners, as these tend to be very frequent but do not provide any interesting semantic information for use in QA. Second, derivation and compounding supersede inflection. For instance, in the QA pair presented in Figure 1 there are two morphological steps between the noun *Australie* (eng:

[1] Co-authors of the present article.

	Quæro	EQueR-Medical	CONIQUE
#Questions	350	200	201
#QA pairs	566	394	664
Avg. question length	8.8	9.9	11.4
Avg. answer length	38.5	29.0	92.4

Table 1: Annotation corpora statistics

	Inflection		Derivation		Compounding	
Corpus (qa pairs)	nbr	%	nbr	%	nbr	%
Conique (664)	159	41.8	188	**49.5**	33	8.7
Quæro (566)	136	**61.8**	80	36.4	4	1.8
EQueR (394)	69	26.4	81	31.0	111	**42.5**

Table 2: Inflection, derivation and compounding in the three corpora

Australia) in the question and the feminine adjective *australienne* (eng: australian) in the answer: the first step is the derivation of the adjective *australien* (eng: australian) out of the noun, and the second one is the inflection of the derived adjective in a feminine form. But the relevant morphological relation between the question and the answer is the derivation of the adjective *australien* out of the noun *Australie*, so that only this one has been annotated. Finally, a specific tag "other" was used to label words that are not directly related (i.e. that are related by more than one morphological process).

Q: Quelle est la capitale de l' **Australie** ?
A: le territoire sur lequel est située la capitale fédérale **australienne**, Canberra .

Figure 1: Example of QA pair where both derivational and inflectional information are available

3. Analysis of the annotated data

At the end of the annotation step, we obtained a set of morphologically related words, that can be studied according to different points of view. First we studied the repartition of morphological relation types such as inflection, derivation and compounding in the three corpora. Then, we analysed in more details the part-of-speech involved in each morphological relation, the grammatical features expressed by the inflectional processes and the semantic types of derivational processes.

3.1. Morphological relation types

	Adjectives		Nouns		Verbs	
	nbr	%	nbr	%	nbr	%
Conique (159)	45	28.3	43	27.0	71	**44.7**
Quæro (136)	9	6.6	55	40.5	72	**52.9**
EQueR (69)	22	31.9	33	**47.8**	14	20.3

Table 3: Parts of speech involved in inflectional processes

The results of the annotation of each corpus according to the different types of morphological relations are presented in Table 2. Each question-answer pair (qa pair) does not necessarily contain a morphological relation, and, more importantly, several pairs of words in the same question-answer pair can be morphologically related to one another, with different morphological relations.

The figures in Table 2 show that each corpus seems to favour a particular type of morphological relation: the Conique corpus contains a majority of derivational relations, while the Quæro corpus comprises more inflectional morphology. As for the EQueR corpus, it presents more compounding than any other kind of morphological relation. Moreover, if we consider the type of morphological relation depending on the corpus, inflection has the greatest proportion in the Quæro corpus, derivation is proportionally more present in the Conique corpus, while compounding is almost absent from Conique and Quæro and very important in EQueR.

The Conique and Quæro corpora show little difference with respect to the proportion of compounding. However, Conique contains more derivational relations than Quæro does. This is due to the way the Conique corpus has been built. It is not based on the output of a QA system, but the answers have been manually retrieved and annotated. QA systems usually have difficulties in dealing with derivational morphology. Moreover, there is a large variation in question and answer length between both corpora, as shown in Table 1. This could also explain the presence of more derivationally related pairs of words in Conique, because the longer the questions and the answers, the more opportunities to observe a derived word and its base. As for EQueR, the great proportion of compounds is certainly related to domain of the corpus: it contains a lot of medical terms, which are often compound nouns, as shown in Figure 2. These morphological characteristics of medical data have already been pointed out by Namer and Zweigenbaum (2004).

In the remainder of this section we detail the annotation results for inflection and derivation only, since there are no morphological resources devoted to compounding which could be evaluted.

	direct relation		two steps		two complex	
	nbr	%	nbr	%	nbr	%
Conique (188)	174	92.6	2	1.0	12	6.4
Quæro (80)	70	87.5	1	1.3	9	11.2
EQueR (81)	70	86.4	3	3.7	8	9.9

Table 4: Derivational steps between the word in the question and the word in the anwser

	Conique (174)		Quæro (70)		EQueR (70)	
	nbr	%	nbr	%	nbr	%
Noun > Adj	41	23.6	16	22.9	**28**	**40.0**
Proper N > Adj	**45**	**25.9**	8	11.4	1	1.4
Noun > Noun	29	16.7	5	7.1	7	10.0
Proper N > N	6	3.4	8	11.4	2	2.9
Adj > Noun	3	1.7	0	0	4	5.7
Verb > Noun	41	23.6	**30**	**42.9**	25	35.7
Other	9	5.1	3	4.3	3	4.3

Table 5: Derivational processes in question-answer pairs

Q: Quelle est la conséquence de la **corticothérapie** sur l' os?
A: Le problème essentiel des **corticoïdes** réside dans leurs effets secondaires (... ostéoporose, ostéonécrose aseptique des têtes fémorales ou parfois humérales ...).

Figure 2: Example question-answer pair from EQueR

3.2. Inflection

The analysis of the inflectional relations found in the three corpora confirms the difference, already observed at the relation type level (Section 3.1.), between Conique and Quæro on the one hand and EQueR on the other hand. Indeed, in Conique and Quæro most inflectional relations are verbal, whereas in EQueR most of them are nominal and the verbal ones are very few, as shown in Table 3.

3.3. Derivation

As shown in Table 2, derivation is important in the three corpora (between 30% and 50% of the pairs). In some cases the word in the question and the word in the answer are morphologically related by more than one derivational step. For instance *lune* (eng: "moon") and *alunissage* (eng: "landing on the moon") or *lait* (eng: "milk") and *allaitement* (eng: "breastfeeding"). In these cases one word is more complex than the other, but the complex word is not directly derived from the less complex. In some other cases, like *joueur* (eng: "player") and *jouable* (eng: "playable") the word in the question and the word in the answer are morphologically related but neither derives from the other. Instead, they both derive from another word, which is *jouer* (eng: "play") for *joueur* and *jouable*. Table 4 shows the proportion of direct derivational relations, non direct derivational relations and cases where both words are complex and derive from another word. The figures show that most derivational relations between a word in the question and a word in the answer are direct (between 86% and 92%). Only very few relations are non direct. There is little to say about the case when the derivational relation is non direct, since in that case the relation between the two words is pretty unpredictable. That is why we focus our study on the pairs which contain one base and one derivative, with only one derivational process between the two.

While focusing on the direct derivational relations, we can evaluate the proportion of different derivational processes involved. Table 5 presents the result of this evaluation. The figures in Table 5 show that the corpora differ with respect to the derivational processes used. While Conique shows more denominal adjectives (about 47% of the derivational processes), Quæro and EQueR seem to favor noun formation processes (with respectively 61% and 54% of the derivational processes). These two particular derivational processes are described below.

3.3.1. Denominal adjectives

In our data, adjectives which derive from a proper noun (Proper N) are always relational adjectives, like *chilien* (eng: "chilean") derived from *Chili* (eng: "Chile"), or *africain* (eng: "african") derived from *Afrique* (eng: "Africa"). Adjectives deriving from a common noun are mostly relational adjectives too, as shown by the figures in Table 6. For instance *présidentiel* (eng: "presidential") derived from *président* (eng: "president"), or *solaire* (eng: "solar") derived from *soleil* (eng: "sun"). However there also are some qualifying adjectives. For instance *âgé* (eng: "old") which derives from *âge* (eng: "age") with the meaning 'having a certain age' or *montagneux* (eng: "mountainous") derived from *montagne* (eng: "mountain") with the meaning 'full of mountains'. Table 6 presents the proportion of relational or qualifying adjectives in our corpora, and shows that relational adjectives are much more frequent in the three corpora. It is also worth noting that the highest proportion of relational adjectives is found in the medical corpus, which confirms previous works such as (Deléger and Cartoni, 2010).

	Relational Adj.		Qualifying Adj.	
	nbr	%	nbr	%
Conique (41)	23	56.1	18	43.9
Quæro (16)	10	62.5	6	37.5
EQueR (28)	24	85.7	4	14.3

Table 6: Types of denominal adjectives

3.3.2. Noun formation processes

As regards the noun formation processes, the three corpora favour deverbal nominalisations, but deadjectival and denominal nominalisations are also found.[2] The formations of noun out of a noun are very few, except in Conique. Those are mostly masculine and feminine profession names, like *infirmier* and *infirmière* (eng: "male/female nurse"), *directeur* and *directrice* (eng: "male/female director"), *président* and *présidente* (eng: "male/female president"), which we considered to be two distinct words rather than one and the same word inflected for gender. There are some suffixed diminutive nouns too, like *rame* (eng: "ream") > *ramette* (eng: "small ream"), and prefixed nouns like *président* (eng: "president") > *vice-président* (eng: "vice-president"). We also considered the formation of a noun out of a proper noun to be a denominal nominalisation. These derived nouns are mostly demonyms (names for the resident of a place) which derive from a location denoting proper noun, like *Colombien* (eng: "Colombian") derived from the country name *Colombie* (eng: "Colombia"). This kind of nouns is found in the Conique and the Quæro corpora, but there are only two in the EQueR corpus, which is not surprising since it is a medical corpus.

Deadjectival nouns are very few in the three corpora. None of them is found in Quæro, and there are just a few of them in the other two corpora. These deadjectival nouns are property nouns, like *toxicité* (eng: "toxicity") which derives from the adjective *toxique* (eng: "toxic"). Not surprisingly deadjectival nouns denoting a property are mostly found in the EQueR corpus. It can be explained by the fact that the medical corpus contains a lot of disease or trouble nouns (like *toxicité* or *insuffisance* "insufficiency") which often refer to the property of being in a particular state (*toxicité* ≈ 'property of being toxic', *insuffisance* ≈ 'property of being insufficient').

As for deverbal nominalisations, they are most often event nouns in the three corpora, like *débarquement* (eng: "landing") derived from the verb *débarquer* (eng: "to land"). Event denoting nouns represent almost 85% of the deverbal nouns, as shown in Table 7. However, there also are a small number of agent nouns in the Conique and the Quæro corpora, but none in the EQueR corpus. For instance *réalisateur* (eng: "director") from *réaliser* (eng: "to direct"). And there is a small set of result nouns like *produit* (eng: "product") which derives from the verb *produire* (eng: "to produce").

[2]The type of nominalisation (deverbal, deadjectival or denominal) depends on the part-of-speech category of the base (verb, adjective or noun, respectively).

3.3.3. Other derivational processes

Other derivational processes include for instance adverbs formation out of adjectives, like *complètement* (eng: "completely") derived from *complet* (eng: "complete"), or *directement* (eng: "directly") derived from *direct* (eng: "direct"). There also are some prefixed deverbal verbs like *déboucher* (eng: "unblock") out of *boucher* (eng: "block") or denominal adjectives like *international* (eng: "international") derived from *nation* (eng: "nation"). Interestingly we observed no deadjectival verb formation (like *national* "national" > *nationaliser* "nationalize") and almost no denominal verb formation. Only four denominal verbs were found in Conique, and none in the other corpora. The absence of denominal verbs could be explained by the rather unpredictable semantic relation between a noun and a derived verb. As stated by Hopper and Thompson (1984) there is an asymmetry in the lexical categories, since a nominalisation still names the event denoted by the verb, whereas a verbalization does not refer to the entity denoted by the noun, but denotes an event associated with that entity. For instance, the noun *destruction* denotes the same event as its base verb *destruct*. But in the case of a denominal verb like *hospitalize*, the verb does not refer to the object denoted by the base noun *hospital*, but denotes some event related to that object. What is more, the events we could associate to an entity are numerous and various. So, the semantic relation between a noun and its derived verb is less informative than that of a verb and its derived noun. It is not surprising then that so few nouns related to their derived verbs were found in the corpora.

4. Evaluation of morphological resources

The set of morphologically annotated data presented in the previous section forms a gold-standard of morphological relations on which we can evaluate the coverage of existing morphological resources.

4.1. Description of the resources

Several resources are available to deal with French morphology. However none of them handles the whole morphology for French. Instead, there are different resources, each of them dealing with a specific area of French morphology. Thus, we took the morphological resources dealing with each type of morphological process we found in the corpora and evaluated them according to their morphological specificity. For inflectional morphology we evaluated two resources : Morphalou and Le*fff*. For derivational morphology we evaluated three different resources : Verbaction and Dubois for deverbal nouns, and Prolexbase for denominal adjectives.

4.1.1. Morphalou

Morphalou is an inflectional lexicon for French.[3] It contains 539,413 inflected forms corresponding to 68,075 lemmas.

[3]http://www.cnrtl.fr/lexiques/morphalou/

	Conique (41)		Quæro (30)		EQueR (25)	
	nbr	%	nbr	%	nbr	%
Verb > Event N	**34**	**82.9**	**25**	**83.3**	**22**	**88**
Verb > Agent N	4	9.8	4	13.3	0	0
Verb > Other N	3	7.3	1	3.4	3	12

Table 7: Semantic types of deverbal nouns in question-answer pairs

4.1.2. Le*fff*

Le*fff* is a syntactic and morphological lexicon for French (Sagot, 2010).[4] It contains morpho-syntactic information for each inflected form, suchs as part of speech, lemma and sub-categorization. Overall it contains 534,763 infected forms.

4.1.3. Verbaction

Verbaction is a lexicon of French action nouns linked to their corresponding verbs (Hathout et al., 2002; Hathout and Tanguy, 2002,).[5] It totals 9,393 verb-noun pairs.

4.1.4. Dubois

This XML resource is based on the description of French verbs by Dubois and Dubois-Charlier (1997).[6] It classifies verbs in semantic and syntactical classes and also provides information about some derivatives of the verbs. Overall it contains 25,609 verb entries and mentions 33,955 derivatives.

4.1.5. Prolexbase

Prolexbase is a multilingual dictionary of proper nouns (Bouchou and Maurel, 2008; Tran and Maurel, 2006).[7] While not targeted at morphology, it nevertheless provides information about relational nouns and adjectives associated with proper nouns, e.g. *Français* and *français* (eng: "French") are explicitly associated with *France*. In some cases, relational nouns and adjectives are not morphologically related to the proper noun, e.g. *britannique* (eng: "british") with *Royaume-Uni* (eng: "United Kingdom"). Overall, it comprises 76,118 lemma and 20,614 derivational relations.

4.2. Evaluation results

4.2.1. Inflection

Two resources, Morphalou and Le*fff*, have been evaluated regarding the inflectional phenomena. Both resources contain approximately the same amount of inflected forms (see previous section), but have been built using different methods. Part of the information in Le*fff* has been automatically acquired and manually validated, while Morphalou's data originate from the TLFNome, the nomenclature of the TLF (Trésor de la Langue Française). In order to evaluate the coverage of the resources, each member of the inflectionally related word pairs identified in our corpus study was looked up in the lexicon. If correctly analysed, both members of the pairs should have the same lemma, and the link between them can be computed. The coverage of each resource was calculated by considering pairs that were correctly analysed, i.e. pairs of words with the same lemma. Table 8 presents the result of the evaluation of Le*fff* and Morphalou for inflectional pairs.

Both resources have very high coverage of inflectional processes in the three corpora. Le*fff* appears to be a little more complete than Morphalou, since its coverage is slightly better in the Conique corpora. Moreover, on the EQueR dataset they differ in the word pairs they cover although they have the same global coverage. Indeed, both of them cover 65 pairs of inflected words out of 69, but the covered pairs are not exactly the same. So that the global coverage made by at least one resource is slightly better than the coverage of one and only one resource. This fact shows that using two different resources for the same type of morphological phenomena can improve the global coverage of the data.

4.2.2. Derivation

Assessing derivational resources is not as straightforward as inflectional ones. The three considered morphological resources that are available for French derivational morphology are designed to address specific morphological phenomena. Dubois and VerbAction contain exclusively deverbal morphology, while Prolexbase only contains demonym nouns and relational adjectives. Consequently, assessing the relevance of these resources can only be done with the appropriate sub-part of the gold-standard. The coverage of VerbAction and Prolexbase was calculated by counting the number of pairs that have been found in them. As for Dubois, it does not literally contain verbal derivatives. Those are only mentioned with specific information from which we can deduce the derivatives. Thus, in order to evaluate the coverage of Dubois we only took into account cases where the derivatives would be automatically computable from information provided in the resource.

As regards the deverbal nouns, Table 9 summarises the coverage of VerbAction and Dubois for event nouns. As we can see, VerbAction has a better coverage than Dubois, especially in lay corpora (Conique and Quæro). As for the deverbal agentive nouns, Dubois covers 100% of the Conique corpus and 75% of the Quæro corpus (no agentive noun has been found in EQueR corpus), while VerbAction does not contain any of them, since it is devoted to action nouns.

As for the demonyms and relational adjectives derived from geographical names, the result of the evaluation of Pro-

[4] http://alpage.inria.fr/~sagot/lefff.html

[5] http://redac.univ-tlse2.fr/lexiques/verbaction.html

[6] http://rali.iro.umontreal.ca/Dubois/

[7] http://www.cnrtl.fr/lexiques/prolex/

Corpus (nbr.)	Lefff		Morphalou		Global coverage	
	nbr.	%	nbr.	%	nbr.	%
Conique (159)	159	100.0	157	98.7	159	100.0
Quæro (136)	135	99.3	135	99.3	135	99.3
EQueR (69)	65	94.2	65	94.2	66	95.6
Total (364)	**359**	**98.6**	**357**	**98.1**	**360**	**98.9**

Table 8: Coverage of inflection

Corpus (nbr.)	VerbAction		Dubois	
	nbr.	%	nbr.	%
Conique (34)	33	97.1	19	55.9
Quæro (25)	25	100.0	9	36.0
EQueR (22)	22	100.0	19	86.4
Total (81)	**80**	**98.8**	**47**	**58.0**

Table 9: Coverage of resources for deverbal event nouns

lexbase is presented in Table 10. We distinguished between Demonym, Relational adjective, and LocOrg (grouping name of place and institutional entities). The figures show that Prolexbase has a very good coverage for both Demonyms derived from a Location name, and relational adjectives derived from Demonyms or Location names. In the Quæro corpus no Demonym>RelAdj pair has been found, and in the EQueR corpus, only one pair LocOrg>RelAdj has been found and is correctly analysed in Prolexbase.

When evaluating the three different types of derivational resources (VerbAction, Dubois and Prolexbase) on the whole gold-standard the coverage is not as high as on specific parts of the gold-standard. Indeed, the global coverage of the three resources is only slightly higher than 50%, as shown in Table 11. Morphological resources are efficient for specific morphological processes. But very frequent phenomena seem to be lacking in the assessed resources, like deverbal agent nouns formation[8] and denominal adjectives formation. This is highly regrettable since the latter process is the second most frequent in the pairs in Conique and Quæro, and the first more frequent in EQueR, as shown in section 3. Consequently, efforts on building resources should be put on this particular phenomenon to address such a frequent issue.

5. Conclusion and Perspectives

In this paper, we have presented an in-depth analysis of the role of morphology in one specific NLP task: Question Answering. Based on a large-scale annotation of three distinct corpora of question-answer pairs, we have built a gold-standard of morphologically related words in question-answer pairs. This gold-standard provides interesting insights on the kind of morphological relations that are mostly implied, and it uncovers those which could have a significant impact on the application performance. Moreover, based on this gold-standard, we have evaluated the coverage of existing morphological resources for French. This evaluation proved that the analyzed French inflectional and derivational resources have a good coverage of the morphological knowledge they target. But some important morphological phenomena are lacking a dedicated resource such as denominal adjectives and agent deverbal nouns. In the future, we hence plan to develop some new French morphological resources for these two phenomena.

[8]Dubois does contain information about deverbal agent nouns. However, these nouns are not explicitly part of the resource and would have to be automatically computed from the indications provided in the resource.

Acknowledgments

This work has been partially financed by OSEO under the Quaero program.

6. References

Christelle Ayache, Brigitte Grau, and Anne Vilnat. 2006. EQueR: the French Evaluation campaign of Question-Answering Systems. In *Proceedings of LREC 2006*.

Matthew W. Bilotti, Boris Katz, and Jimmy Lin. 2004. What Works Better for Question Answering: Stemming or Morphological Query Expansion. In *Proceedings of the Information Retrieval for Question Answering (IR4QA) Workshop at SIGIR 2004*, Sheffield, England.

Béatrice Bouchou and Denis Maurel. 2008. Prolexbase et LMF: vers un standard pour les ressources lexicales sur les noms propres. *Traitement Automatique des Langues*, 49(1):61–88.

Mathias Creutz, Teemu Hirsimäki, Mikko Kurimo, Antti Puurula, Janne Pylkkönen, Vesa Siivola, Matti Varjokallio, Ebru Arisoy, Murat Saraçlar, and Andreas Stolcke. 2007. Morph-based speech recognition and modeling of out-of-vocabulary words across languages. *ACM Transactions on Speech and Language Processing (TSLP)*, 5(1):1–29.

Louise Deléger and Bruno Cartoni. 2010. Adjectifs relationnels et langue de spécialité : vérification d'une hypothèse linguistique en corpus comparable médical. In *Proceedings of TALN 2010*.

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

Corpus	Morphological relation (nbr.)		Found in Prolexbase	
			nbr.	%
Conique	Demonym - Rel Adj	(1)	1	100.0
	LocOrg - Demonym	(6)	6	100.0
	LocOrg - Rel Adj	(45)	43	95.6
Quæro	LocOrg - Demonym	(8)	5	62.5
	LocOrg - Rel Adj	(8)	8	100.0
EQueR	LocOrg - Rel Adj	(1)	1	100.0
Total		**69**	**64**	**92.7**

Table 10: Coverage of Prolexbase for Geographic morphological relation

Corpus (nbr.)	Global coverage	
	nbr.	%
Conique (174)	98	56.3
Quæro (70)	41	58.6
EQueR (70)	26	37.1
Total (314)	**165**	**52.5**

Table 11: Global coverage of the three derivational resources on derivational pairs

Michael Fuller and Justin Zobel. 1998. Conflation-based comparison of stemming algorithms. In *Proceedings of the Third Australian Document Computing Symposium*, pages 8–13, Sydney.

Ulrich Germann. 2008. Yawat: yet another word alignment tool. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies (HLT '08)*, pages 20–23.

Arnaud Grappy, Brigitte Grau, Olivier Ferret, Cyril Grouin, Véronique Moriceau, Isabelle Robba, Xavier Tannier, Anne Vilnat, and Vincent Barbier. 2010. A Corpus for Studying Full Answer Justification. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Udo Hahn, Martin Honeck, and Stefan Shulz. 2003. Subword-Based Text Retrieval. In *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*, Big Island, Hawaii, January 06 - 09.

Donna Harman. 1991. How effective is suffixing? *Journal of the American Society of Information Science*, 42(1):7–15.

Nabil Hathout and Ludovic Tanguy. 2002. Webaffix: Discovering Morphological Links on the WWW. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pages 1799–1804, Las Palmas de Gran Canaria, Espagne. ELRA.

Nabil Hathout, Fiammetta Namer, and Georgette Dal, 2002. *Many Morphologies*, chapter An Experimental Constructional Database : The MorTAL Project, pages 178–209. Cascadilla Press.

P.J. Hopper and S.A. Thompson. 1984. The discourse basis for lexical categories in universal grammar. *Language*, 60:703–752.

Philipp Koehn and Hieu Hoang. 2007. Factored Translation Models. In *Proceedings of EMNLP-CoNLL 2007*, pages 868–876, Prague, Czech Republic.

Martin Lennon, David S. Pierce, Brian D. Tarry, and Peter Willett. 1988. An evaluation of some conflation algorithms for information retrieval. *Journal of Information Science*, 3(4):177–183.

Paul McNamee, Charles Nicholas, and James Mayfield. 2009. Addressing morphological variation in alphabetic languages. In *SIGIR '09: Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 75–82, New York, NY, USA. ACM.

Fabienne Moreau and Vincent Claveau. 2006. Extension de requêtes par relations morphologiques acquises automatiquement. In *Actes de la Troisième Conférence en Recherche d'Informations et Applications CORIA 2006*, pages 181–192.

Fiammetta Namer and Pierre Zweigenbaum. 2004. Acquiring meaning for french medical terminology: contribution of morphosemantics. *Eleventh MEDINFO International Conference*, pages 535–539.

Ludovic Quintard, Olivier Galibert, Gilles Adda, Brigitte Grau, Dominique Laurent, Véronique Moriceau, Sophie Rosset, Xavier Tannier, and Anne Vilnat. 2010. Question Answering on Web Data: The QA Evaluation in Quæro. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Benoît Sagot. 2010. The Lefff, a Freely Available and Large-coverage Morphological and Syntactic Lexicon for French. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Mickaël Tran and Denis Maurel. 2006. Prolexbase : un dictionnaire relationnel multilingue de noms propres. *Traitement Automatique des Langues*, 47(1):115–139.