Leveraging RDF Graphs for Crossing Multiple Bilingual Dictionaries

Marta Villegas*, Maite Melero*, Jorge Gracia♣, Núria Bel*
*Universitat Pompeu Fabra,♣ Ontology Engineering Group, Universidad Politécnica de Madrid
E-mail: marta.villegas@upf.edu, maite.melero@upf.edu, nuria.bel@upf.edu jgracia@fi.upm.es

Abstract

The experiments presented here exploit the properties of the Apertium RDF Graph, principally cycle density and nodes' degree, to automatically generate new translation relations between words, and therefore to enrich existing bilingual dictionaries with new entries. Currently, the Apertium RDF Graph includes data from 22 Apertium bilingual dictionaries and constitutes a large unified array of linked lexical entries and translations that are available and accessible on the Web (http://linguistic.linkeddata.es/apertium/). In particular, its graph structure allows for interesting exploitation opportunities, some of which are addressed in this paper.
Two experiments are reported: in the first one, the original EN-ES translation set was removed from the Apertium RDF Graph and a new EN-ES version was generated. The results were compared against the previously removed EN-ES data and against the Concise Oxford Spanish Dictionary. In the second experiment, a new non-existent EN-FR translation set was generated. In this case the results were compared against a converted wiktionary English-French file.
The results we got are really good and perform well for the extreme case of correlated polysemy. This led us to address the possibility to use cycles and nodes degree to identify potential oddities in the source data. If cycle density proves efficient when considering potential targets, we can assume that in dense graphs nodes with low degree may indicate potential errors.

Keywords: Apertium,, LLOD, RDF/OWL, graphs, multilingual lexica, bilingual lexica, automatic lexical aquisition

1. Introduction

The Apertium RDF Graph contains the RDF version of the Apertium [1] bilingual dictionaries (Forcada et al., 2011), which have been transformed into RDF and published on the Web following the Linked Data principles. As described by Gracia et al.. (2015), the core linguistic data of the Apertium RDF Graph was modeled using *lemon*, the LExicon Model for ONtologies (McCrae et al., 2012) while the translations between lexical entries used the *lemon translation module* (Gracia et al., 2014).
Currently, the Apertium RDF Graph includes data from 22 Apertium bilingual dictionaries and it is expected that more Apertium data will be included in the near future. As result of the generation of the Apertium RDF Graph, a large unified array of linked lexical entries and translations is available and accessible on the Web (http://linguistic.linkeddata.es/apertium/). Its graph structure allows for interesting exploitation opportunities, some of them addressed in this paper. In particular, we propose a method to discover candidate translations for a given lexical entry and an algorithm to compute the confidence degree for such translations, based on the density exploration of the graph.
The rest of the paper is organized as follows. In Section 2, related work on deriving bilingual lexical information from existent resources is discussed. Section 3 introduces our proposal for inferring indirect translations based on the exploration of the graph's cycle density. The algorithm to compute a score for the cycles is presented in Section 4. Section 5 discusses some illustrative examples. Section 6 details the experiments performed. The discussion of the obtained results is in Section 7 and, finally, conclusions and future work can be found in Section 8.

[1] https://www.apertium.org

2. Related work

Deriving new bilingual lexica from already existing ones in not new. Initial proposals typically used a pivot language to derive a new bilingual lexicon between two Source and Target languages, provided that the pairs Source/Pivot and Target/Pivot were already available. When using a pivot language to construct a bilingual dictionary, it is mandatory to discriminate inappropriate equivalences between words caused by translation ambiguities. A method to identify such incorrect translations was proposed by Tanaka & Umemura (1994) when constructing bilingual dictionaries intermediated by a third language. The method, known as one time inverse consultation (OTIC), was adapted by Lim et al. (2011) in the creation of multilingual lexicons from bilingual lists of words.
More recently, different algorithms exploiting graph properties have been proposed to derive, enrich and/or validate lexical resources. Graph algorithms allow working with a larger number of lexicons. For instance the SenseUniformPaths algorithm (Soderland, 2009), based on graph sampling, uses probabilistic methods to infer lexical translations. The SenseUniformPaths was used in the generation of the PANDICTIONARY. According to this algorithm, all nodes on a translation circuit share a sense with high probability, unless there is a correlated polysemy among the nodes on the path (that is a pair of nodes sharing the same polysemy). To avoid correlated sense-shifts the SenseUniformPaths algorithm identifies (and prunes) 'ambiguous cycles', i.e. cycles with sets of nodes sharing multiple senses. Flati et al. (2013) introduce the notion of cyclic and quasi-cyclic graph paths. They use it for the disambiguation and validation of bilingual dictionaries, and to automatically identify synonyms aligned across languages.
The approach presented in this paper is grounded on Soderland's in that we use cycles to identify potential targets. It differs in that our source data (i.e. the Apertium

dictionaries) has no notion of sense and, more importantly, we use nodes' degree and graph density (more specifically, cycle density) to rate confidence value. Once a cycle C is identified, our algorithm does not need to identify potential 'ambiguous cycles' but rather relies on graph properties, which is computationally less expensive. Note in addition that identifying potential ambiguous cycles only works well provided that source dictionaries are complete (all translations for a given word are encoded) and have similar coverage, which is not the case for the Apertium RDF graph, as the original dictionaries are incomplete and quite unbalanced.

3. Computation based on cycle density

Word polysemy is a linguistic feature that prevents to consider translation as a transitive relation. For instance, by knowing that the Spanish translation of the English word *wrist* is *muñeca* and that *muñeca* translates into French *poupée* we would be wrong concluding that *wrist*-en→*poupée*-fr. More generally, given a chain such as a→b→c→d→e, we cannot assume that every node is reachable from every other node (having a complete graph).

Following Soderland, we base our experiment in cycles (instead of simple chains). A cycle is a sequence of nodes (vertices) starting and ending in the same node with no repetitions of vertices and edges allowed. We assume that the probability that a cycle (in our case: a cycle of translations) becomes a complete graph is higher than the probability that a simple translations path becomes complete. Thus, the probability of having *wrist*-en→ *poignet*-fr is much higher in a cycle like the one in (a) than the probability of having *wrist*-en→*poupée*-fr in a path like that in (b):

(a) *wrist*-en → *muñeca*-es → *poignet*-f r →*canell*-ca

(b) *wrist*-en → *muñeca*-es → *poupée*-f r →*nina*-ca

In addition, we assume that the density of a cycle is crucial to calculate the probability that two nodes with no direct connection be good translation candidates. The density of a graph depends on the number of vertices and edges on that graph. Thus, the higher the number of edges a graph has the higher its density is. Density is defined as D=|E|/|V|*(|V|-1) where E is the number of edges and V is the number of vertices in the graph. The minimal density is 0 and the maximal density is 1 (for complete graphs). The table below shows three example graphs with different densities. Whereas they all have the same number of vertices, they differ in the number of edges and therefore in their density:

(A ↔ B ↔ C ↔ D ↔ A)	D = 8 / 4*3 = 0.66
(A ↔ B ↔ C ↔ D↔ A) + (B↔ D)	D = 10/4*3 = 0.83
(A ↔ B ↔ C ↔ D↔ A) + (B↔ D)+ (A ↔ C)	D = 12/4*3 = 1

In this example, we assume that the probability of A↔C is higher in the second graph (with D=0.83) than in the first one (D=0.66) where no other edges, except those forming the main cycle do exist.

To avoid potential translation errors arising from polysemy, we impose the following restriction on translation cycles: Cycles starting and ending in word W of language L may not contain other words from L. In our initial experiments we did not introduce such a restriction. We found that, in most cases, when two nodes in the cycle belong to the same language, they were synonyms (which is what one would expect). Note, however, that source data is not error free (possibly because some Apertium dictionaries were partially automatically created using transitive relations) and such reentrances may produce false results. The example below illustrates such a situation:

poignet-fr→*pojno*-eo→*poupée*-fr→*doll*-en→ *muñeca*-es

In this example, the Esperanto word *pojno* (wrongly) links to French *poupée*, and this initiates a different 'sense path'. Since, the Spanish word *muñeca* is polysemous (meaning both *doll* and *wrist*), the path ends back to *poignet*, (FR) wrongly suggesting that *poignet* and *poupée* may be synonyms in French.

Note that, such a restriction does not prevent us from having cycles with 'correlated polysemy'. In the example below, we evaluate the Catalan word *canell* (meaning *wrist)*, as potential target of *doll*. *Canell* is introduced by *pojno* and leads to the Spanish polysemous word *muñeca*.

> ***doll-en / canell-ca*** **score 0.6**
> **Best cycle**: *doll-en* → *nina-**ca*** → *poupée-fr* → *pojno-eo* → *canell-**ca*** → *muñeca-es* → *doll-en*

Note that, eventhough *pojno* and *muñeca* introduce a correlated polysemy, the algorithm still gives a 'low' confidence score (0.6) to the potential target *canell*. We get the same low score when evaluating the French word *poignet* as a potential target for *doll*:

> ***doll-en / poignet-fr*** **score 0.6**
> **Best cycle**: *doll*-en → *nina*-ca → *poupée*-**fr** → *pojno*-eo → *poignet*-**fr** → *muñeca*-es → *doll*-en

As we will see in the next section, we may impose a further restriction on cycles and reject those where two or more nodes belong to the same languages. In this case, the cycles above (those involving *canell*-ca and *poignet*-fr) would not be considered.

4. Initial algorithm

In order to be able to find all cycles involving a root word *W* in such a huge graph as the Apertium RDF Graph, we need to reduce this into a manageable sub-graph we call the context of *W*. The context of *W* is defined as including:

1) all translations of *W* in any language, trans(*W*)
2) for each element in trans(*W*), all its translations trans(trans(*W*))
3) for each element in trans(trans(*W*)), all its translations trans(trans(trans(*W*)))

Context(*W*) = *W* + trans(*W*) + trans(trans(*W*)) + trans(trans(trans(W)))

This gives a list of "source/target" pairs, like *doll/nina, muñeca/poupé, muñeca-pojno* etc that constitutes the context of W (*doll*).

Once the Context of *W* is defined, we can compute the **cycles of *W*** (cycles(W)) occurring in that Context. Namely, those paths starting and ending in *W* with no node repetitions. When computing cycles, we limit their length to avoid over computation, thus whenever a path reaches 7 nodes (or 6 to speed up experiments) and no cycle has been found the path is rejected. Optionally we can remove cycles containing nodes with repeated languages (disallowing the cycles we discussed before involving *poignet*-fr and *canell*-ca).

Having cycles(*W*) allows identifying the **potential targets** of *W*, that is: those words in the cycles(*W*) which are not directly linked to *W*. Note that words in the Context(W) which are not in a cycle are discharged as potential targets. This means that nodes beyond bridges are 'dismissed' (a bridge is an edge whose removal renders the graph disconnected). The dimension of Context(W) and the number of cycles it contains, varies from word to word. Thus we find large contexts such as that for *boy* (with 194 nodes and 539 vertices) and rather small contexts as that of *veterinarian* (with 15 nodes and 31 vertices).

Finally, for each potential target T, we get the cycles containing T and calculate their density (where density is the ration between vertices ad edges: D = V / N*(N-1). We use the more dense cycle to assign the **confidence score** so that T be an admissible translation of W. In the example below, we give the results for *'doll - poupée'*, in the 'no language repetition' mode:

> ***doll -en* / *poupée-fr*** **score 0.83333**
> **Best cycle***: doll-en → pupo-eo → poupée-fr' → nina-ca → doll-en*

As we will see in Section 6, cycle's density alone is not enough to discriminate wrong targets introduced by correlated polysemy.

5. Some illustrative examples

In Table 1 we show some examples involving different root words, namely: *doll*, *wrist*, *poignet*, *rede* and *bambino*. For each root word we give: (i) the number of different words in its context (nodes), (ii) the number of translation pairs in the context (edges), (iii) the number and the list of already known targets in the Apertium data and (iv) the number and list of potential targets found, together with the confidence score. In these examples, we run the experiment allowing for language repetition in cycles. Since we do not want to introduce wrong translations, we opt for higher precision at the expense of recall, and, therefore, defined a threshold of 0.7.

In the *doll* example we correctly get a higher score for *poupée* (0.833) than to *poignet* or *canell* that fall below the threshold (0.6). We cannot avoid the 0.7 score to *pojno* because of the incorrect links to *poupée* in the source data, as mentioned above. In the *wrist* example, correlated polysemy wrongly produces *poupée* together with the correct *poignet*. In the *poignet* example we get 4 correct translations while getting lower scores for wrong "doll sense" targets (i.e. *doll*, *nina*, *boneca* and *pipa*). In next section we address again these 'extreme' examples.

The Portuguese *rede* example (meaning *net*) demonstrates that when correlated polysemy is not involved results are expectedly good, with both high precision and excellent coverage.

Finally, the *bambino* example, in the last column, shows that even in the case when the root word only has correspondences to one single language, Catalan in this case, the system is able to produce good results. In this example, we get 15 translations out of a context with 83 different words. Notice that, if instead of having two initial translations as in here (*nen* and *xiquet*) we had only one; no cycle would have been created. Note, however, that few good translations are rejected. Finally, notice that running the experiment with the 'no repetition language' option would produce no results.

ROOT	**doll (EN)**	**wrist(EN)**	**poignet (FR)**	**rede(PT)**	**bambino(IT)**
words	58	43	37	41	83
trans. pairs	135	100	89	108	212
known targets	muñeca-es muñeco-es nina-ca moneca-gl boneca-gl pupo-eo	canell-ca muñeca-es pulso-gl moneca-gl manradiko-eo manartiko-eo pojno-eo eskumutur-eu	muñeca-es pojno-eo manumo-eo manartiko-eo manradiko-eo	xarxa-ca xàrxia-ca red-es rede-gl hamaca-es	xiquet-ca nen-ca
Potential targets	canell-ca 0.6 **boneco-pt 0.7** **poupée-fr 0.833** ninot-ca 0.6 monyica-ca 0.6 pipót-oc 0.533 bambola-it 0.6 **pojno-eo 0.7** **pipa-oc 0.833** poignet-fr 0.6	**poupée-fr 0.833** monyica-ca 0.6 **nina-ca 0.7** bambola-it 0.6 pipa-oc 0.6 **poignet-fr 0.833**	**wrist-en 0.833** **eskumutur-eu 0.7** puny-ca 0.666 doll-en 0.6 nina-ca 0.523 **canell-ca 0.833** **moneca-gl 0.7** boneca-gl 0.523 pipa-oc 0.464	**filat-oc 0.833** **ret-oc 0.833** **hialat-oc 0.833** **rete-it 0.7** **sare-eu 0.7** **net-en 0.833** **filet-fr 0.7** **reseau-fr 0.833** **network-en 0.833** **hilat-oc 0.833**	**nene-es 0.7** girl-en 0.666 **menino-pt 0.7** **child-en 0.7** boy-en 0.666 chiquillo-es 0.571 criança-pt 0.6 **mainat-oc 0.7** **niño-es 0.7** **kid-en 0.7**

				malhum-oc 0.833 **xarxa-oc 0.833** **reto-eo 0.833** **fialat-oc 0.833**	**knabo-eo 0.7** peque-es 0.666 **mainatge-oc 0.7** **mainada-oc 0.**715 enfant-fr 0.666

Table 1: Examples with some root words: *doll, wrist, poignet, rede* and *bambino*.

ROOT	**forest-en**		
Words (nodes) in context (88)	bòsc-oc, verre-fr, vidrio-es, vaso-gl, leña-es, fuorto-eo, llenya-ca, búcaro-es, floresta-pt, monte-es, arboleda-es, floreiro-gl, llenyam-ca, ligno-eo, pahar-ro, veire-oc, xerro-gl, bosco-it, vazo-eo, selva-ast, wood-en, sèuva-oc, vaso-es, bosque-pt, viesca-ast, gerro-ca, bosko-eo, jarro-es, vas-ca, kristal-eu, arbareto-eo, fusta-ca, exploitationforestière-fr, beira-eu, vidre-ca, forsto-eo, forest-en, brulligno-eo, bois-fr, foresta-it, copa-ca, selva-pt, bosc-ca, copo-pt, madeira-gl, floresta-es, kornaro-eo, pinar-es, pitxer-ca, fort-fr, abiarbaro-eo, baso-eu, boscatge-ca, jarrón-es, glass-en, zur-eu, got-ca, cristal-es, glaso-eo, bosque-gl, monte-ast, selva-es, kornobranaro-eo, arbreda-ca, fustam-ca, copa-es, fraga-gl, mata-ast, madera-es, copa-gl, pdure-ro, pineda-ca, praarbaro-eo, vase-en, vidro-gl, codru-ro, bosque-es, pinarbaro-eo, cristal-gl, selva-ca, selva-gl, montaña-es, ornamavazo-eo, arbaro-eo, angalo-eo, woodland-en, vitro-eo, vasu-ast		
Edges in context	161		
Known targets (9)	bosc(ca), bosque(es), fraga(gl), bosque(gl), arbaro(eo), abiarbaro(eo), pinarbaro(eo), forsto(eo), baso(eu)		
Potential targets (6)	**bois(fr) 0.900** **fort(fr) 0.900**	**bósc(oc) 0.833** **bosque(pt) 0.833**	**floresta(pt) 0.700** selva(es) 0.619
Dismissed words (73)	exploitationforestière-fr, madeira-gl, angalo-eo, verre-fr, sèuva-oc, vidrio-es, pdure-ro, vaso-es, vaso-gl, floresta-es, jarro-es, leña-es, pineda-ca, vase-en, fuorto-eo, pitxer-ca, cristal-gl, viesca-ast, pinar-es, boscatge-ca, gerro-ca, bosko-eo, glass-en, llenya-ca, vidro-gl, codru-ro, monte-es, búcaro-es, foresta-it, vas-ca, kristal-eu, arbareto-eo, zur-eu, cristal-es, jarrón-es, glaso-eo, selva-pt, arboleda-es, floreiro-gl, llenyam-ca, beira-eu, monte-ast, ligno-eo, selva-gl, copa-gl, vidre-ca, pahar-ro, montaña-es, ornamavazo-eo, madera-es, forest-en, got-ca, vasu-ast, fusta-ca, kornaro-eo, copa-es, kornobranaro-eo, arbreda-ca, brulligno-eo, fustam-ca, veire-oc, xerro-gl, copa-ca, woodland-en, selva-ca, mata-ast, vitro-eo, bosco-it, praarbaro-eo, copo-pt, vazo-eo, selva-ast, wood-en		
Num of cycles	1261		
Cycles with P.T.	1204		
Uniq Lang Cycles	476		

Table 2: "Forest-en" full example

bois-fr	0.9	5	[bosque-es, bosc-ca, bois-fr, arbaro-eo, forest-en]
fort-fr	0.9	5	[bosque-es, fort-fr, bosc-ca, arbaro-eo, forest-en]
bòsc-oc	0.833	4	[bosque-es, bòsc-oc, bosc-ca, forest-en]
bosque-pt	0.833	4	[bosque-gl, bosque-pt, bosque-es, forest-en]
floresta-pt	0.7	5	[fraga-gl, floresta-pt, bosque-gl, bosque-es, forest-en]
selva-es	0.619	7	[bosque-es, bosc-ca, arbaro-eo, fort-fr, selva-es, baso-eu, forest-en]

Table 3: "Forest-en" best targets & cycles (with repeated languages)

bois-fr	0.9	5	[bosque-es, bosc-ca, bois-fr, arbaro-eo, forest-en]
fort-fr	0.9	5	[bosque-es, fort-fr, bosc-ca, arbaro-eo, forest-en]
bosque-pt	0.833	4	[bosque-gl, bosque-pt, bosque-es, forest-en]
bòsc-oc	0.833	4	[bosque-es, bòsc-oc, bosc-ca, forest-en]
selva-es	0.533	6	[bosc-ca, arbaro-eo, fort-fr, selva-es, baso-eu, forest-en]

Table 4: "Forest-en" best targets & cycles (without repeated languages)

In Table 2 we give the complete example for the English word *forest*. This full example shows that the set of context words, which is usually rather large, may include semantically related words, such as *selva*, *pinar* and *madera* together with totally unrelated words introduced by polysemic nodes in the cycle. Thus, in the apparently uncontroversial case of *forest*, the context includes unexpected words such as *vase*, *vaso*, *glass*, *cristal*, etc introduced by the polysemic Basque word *baso* (meaning *forest* and *glass*). Because these words do not occur in any cycle of *forest*, they are not considered as potential targets. It is interesting to see the number of dismissed words, up to 73 and the number of cycles found: 1261 when allowing for language repetitions and 476 when disallowing repetition. Finally, tables 3 and 4 give the scores obtained when running in repetition/non-repetition language modes respectively. Note that, when disallowing language repetition, *floresta*-pt is no longer considered and the Spanish candidate *selva* gets a lower score.

As a final note, we observe that in all cases the results obtained imply a substantial increase of lexical coverage. Table 5 shows the increase gained for each root word from the examples above.

Root	Known tr.	New tr.	Increment
doll -en	6	4	66%
poignet-fr	5	4	80%
rede-pt	5	14	280%
bambino-it	2	9	450%
forest-en	9	5	55,55%

Table 5: Increase in number of translations

It is also interesting to observe the coverage increase in terms of target languages involved.

Root	Lang	New Lang	%
doll -en	es, ca, gl, eo	pt, fr, oc	75%
poignet-fr	es, eo	en, eu ,ca, gl	200%
rede-pt	ca, e, gl	oc, it,eu, en,fr,eo	200%
bambino-it	ca	es, pt, en oc, eo	500%
forest-en	ca, es, gl, eo, eu	fr, oc, pt	55,5%

Table 6: Increase in number of languages

6. Experiments

In order to evaluate our method, two different experiments were performed. In the first one, the original EN-ES translation set was removed from the Apertium RDF Graph and a new EN-ES version was generated. The results were compared against the previously removed EN-ES data and against the "Concise Oxford Spanish Dictionary: Spanish-English/English-Spanish" dictionary (COSD)[2]. In the second experiment a new non-existent EN-FR translation set was generated. In this case the results were compared against a converted wiktionary English-French file[3].

Table 7 gives some figures for the two scenarios. In the EN-ES experiment we had three different translation sets involving English with a total amount of 39839 correspondences (translations). In the EN-FR experiment, we had four different translation sets involving English with a total amount of 52574 correspondences.

EN-ES	EN-FR
14613 EN-CA	14613 EN-CA
16258 EN-EO	16258 EN-EO
8968 EN-GL	12735 EN-ES
	8968 EN-GL
Total: 39839	Total: 52574

Table 7: Number of translations in both experiments

We run the **EN-ES experiment** on the list of 18356 distinct English nouns in the Apertium RDF Graph and got the following results:

Total English nouns tested	18356	100%
Nouns with no Spanish cycle/potential target	12880	70.16%
Nouns with Spanish cycle/potential target	5476	29.83%

Table 8: Results for the EN nouns

The 5476 nouns with a Spanish cycle produced a total amount of 6578 potential targets when running in 'no-language repetition' mode and 7007 when allowing for language repetition. Tables 9 and 10 show the results when testing these candidates against the reference dictionaries. In Table 9 we can see the scores when testing against the Apertium original dataset. Scores for the 'non-language repetition' mode are on the left and scores for 'language repetition' mode on the right. Note that only the 0.08% of suggested targets got a score below 0.5 (for no-repetition mode) and 0.44% in the case of repetition mode.

[2] With 13785 en-es correspondences for nouns.

[3] http://wiki.webz.cz/dict/files/english-french.txt

TARGETS	%	SCORE	TARGETS	%	SCORE
5	0.08%	0.4667	31	0.44%	0.4667
240	3.65%	0.5000	2	0.03%	0.4762
57	0.87%	0.5333	227	3.24%	0.5000
247	3.75%	0.5500	3	0.04%	0.5179
29	0.44%	0.5667	90	1.28%	0.5333
108	1.64%	0.6000	4	0.06%	0.5476
236	3.59%	0.6500	232	3.31%	0.5500
322	4.90%	0.6667	148	2.11%	0.5667
107	1.63%	0.7000	1	0.01%	0.5714
2800	42.57%	0.7500	180	2.57%	0.6000
70	1.06%	0.8333	88	1.26%	0.6333
2357	35.83%	0.9167	234	3.34%	0.6500
6578			371	5.29%	0.6667
			130	1.86%	0.7000
			2836	40.47%	0.7500
			3	0.04%	0.7667
			70	1.00%	0.8333
			2357	33.64%	0.9167
			7007		

Table 9 Testing against the Apertium EN-ES data

Similarly, the results are also very good when validating the candidates against the COSD dictionary. In this case, however, the amount of potential translation pairs found in the reference dictionary is lower: 4902 out of 6578 (74.52%) when running in no-repetition mode and 5030 out of 7007 (71.78%) when allowing for language repetition. It is clear that, in this scenario, cycle computation produced some 'extra' candidates that could not be evaluated against the reference data. We may argue that these 'extra' candidates are wrong candidates as they are not included in the COSD dictionary but the fact that they are all in the original Apertium Translation Set led us to consider them 'good' candidates. In any case, for the potential targets that could be checked against reference data the results prove that nearly all of them got a score above 5.

TARGETS	%	SCORE	TARGETS	%	SCORE
3	0.06%	0.4667	16	0.32%	0.4667
85	1.73%	0.5000	80	1.59%	0.5000
24	0.49%	0.5333	28	0.56%	0.5333
182	3.71%	0.5500	173	3.44%	0.5500
20	0.41%	0.5667	61	1.21%	0.5667
42	0.86%	0.6000	69	1.37%	0.6000
171	3.49%	0.6500	29	0.58%	0.6333
173	3.53%	0.6667	178	3.54%	0.6500
47	0.96%	0.7000	181	3.60%	0.6667
2030	41.41%	0.7500	54	1.07%	0.7000
50	1.02%	0.8333	2050	40.76%	0.7500
2075	42.33%	0.9167	2	0.04%	0.7667
4902			50	0.99%	0.8333
			2075	41.25%	0.9167
			5030		

Table 10 Testing against the COSD dictionary

For the **EN-FR experiment**, we wanted to validate the candidates against the reference data. Since our reference dictionary was a rather small one, we run the experiment on the set of 4824 English nouns in that dataset[4]. As the table below shows, 2112 English words provided some cycle with a French candidate, 1415 words did not provide any French cycle and 1297 words were 'irrelevant' for our purposes (irrelevant words are those which fail to provide any results in any of the 3 successive SPARQL queries we run to set the context of W).[5]

Total English nouns tested	4824	100%
'irrelevant' in Apertium data	1297	26.88%
In Apertium but no 'French' cycle	1415	29.33%
In Apertium with 'French' cycle/potential target	2112	43.78%

These 2112 English nouns produced 2745 potential French targets that were evaluated against the reference dictionary. As shown in table below, 1858 candidates were found in the reference data whereas 887 candidates were not found there.

EN-FR pairs produced by the cycle computation	2545	100%
also found in the wiktionary data	1858	73%
not found in the wiktionary data	887	27%

Table 11 summarizes the scores obtained for the 1858 EN-FR candidate pairs also included in the reference dictionary. Again, all validated candidates (all but 3) got a score above 0.5.

TARGETS	%	SCORE
3	0.16%	0.467
12	0.65%	0.500
11	0.59%	0.533
53	2.85%	0.600
53	2.85%	0.667
83	4.47%	0.700
913	49.14%	0.833
730	39.29%	0.900
1858		

Table 11: EN-FR validation results

Though an initial manual inspection showed that the 'extra' translations (the 27% not included in the reference data) were correct, we checked the *doll/wrist* examples and found that (i) in the 'no language repetition' mode the wrong pair *doll/poignet* was no longer produced but (ii) the *wrist/poupée* pair was produced in both modes (and with a high score). Though we were not able to perform a complete checking, we may conclude that 'extra' targets are correct except when correlated polysemy occurs.
This led us to be more restrictive when accepting candidate cycles by imposing some conditions.

[4] The ones we knew we could check.
[5] For EN-FR experiment differences between no-repetition and repetition mode were so small that we only report the non-repetition figures.

Essentially, we force longer cycles (5 or 6 nodes minimum depending on the contexts) and relay on cycle density, assuming that in dense cycles the probability of identifying new source/target candidates is higher. Note, however, that 'cycle density' evaluates source-target probability without considering source nor target themselves; only the density of the cycle is considered. In correlated polysemy scenarios, 'wrong potential targets' are expected to have less degree than 'right' ones. This led us to include source and target degree into the formula (node degree is the number of edges connected to the node). Basically, we require that at least source or target need to have more than two edges. The eventual calculation goes as follows:

(1) **minimal length of cycles**: for words with small contexts we require at least 5 nodes, for words with big contexts we require a minimum of 6 nodes. Small contexts are those where the root word has up to 5 translations. Big contexts are those in which the root word has more than 5 translations.
(2) when **target has >2 edges**, we calculate density score and get those above 0.5 (nearly all)
(3) **when target has only 2 edges**, we require that source word be linked at least with 50% of the far-nodes' and require a score above 0.7. Far nodes are those in the cycle of W that are not next to W. This means that in a 6 node cycle like: W--N1--TARGET--N2--N3--N4--W all nodes but TARGET are required to be linked to W.

Note that without these restrictions, in the *wrist* example, we got the same results for *poignet* and *poupée*. As we can see below, accepting small cycles (4 nodes) produces wrong candidates when cross polysemy occurs:

score	cycle / potential target
0.833	pojno-eo, **poupée-fr** , muñeca-es, **wrist**
0.833	pojno-eo, **poignet-fr**, muñeca-es, **wrist**

When requiring longer cycles we get lower scores and some candidates may be refused. Now, in the *wrist* example (Table 12), both *poignet* and *poupée* are refused as they only have 2 edges and 0.6 score. Though we miss a good target, at this point, we rather prefer precision than recall. Note here, that in the *poignet* case, one could expect a "3 edge" target (as in the case that *poignet* were correctly linked to *canell* or *monecan*); whereas in the *poupée* case, a "3 edge" target is not possible (as *poupée* is 'out of context'). Note, finally, that *wrist* correctly generates two targets (*nina* and *pipa*) as these have 3 edges (as expectedly, '*doll'* senses are better connected in the *doll* cycle)

score		cycle / potential target /edges
0.666	fail	monecan-gl, muñeca-es, **poignet-fr-(2)**, pojno-eo, canell-ca, **wrist**
0.666	fail	moneca-gl, muñeca-es, **poupée-fr-(2)**, pojno-eo, canell-ca, **wrist**
0.6	OK	pojno-eo, poupée-fr, **nina-ca-(3)**, pipa-oc, muñeca-es, **wrist**
0.6	OK	pojno-eo, poupée-fr, nina-ca, **pipa-oc-(3)**, muñeca-es, **wrist**

Table 12 *Wrist* example

In the *doll* example in Table 13, all candidates are correctly generated as they all have 3 edges (a 'good' connectivity) and a score above 0.5. As we saw in Section 3, *poignet* occurs as potential target in "language repetition mode", but even in this case, *poignet* would be rejected as it only has 2 edges as it occurs in a "doll sense" cycle.

score		cycle / potential target /edges
0.666	OK	pupo-eo, poupée-fr, nina-ca, **boneco-pt-(3)**, muñeco-es, **doll**
0.666	OK	pupo-eo , poupée-fr, nina-ca, **pipa-oc-(3),** muñeca-es, **doll**
0.6	OK	nina-ca , poupée-fr , **pojno-eo-(3),** muñeca-es, moneca-gl, **doll**
0.666	OK	pupo-eo, **poupée-fr-(3),** nina-ca , pipa-oc, muñeca-es, **doll**

Table 13: the *doll* example

Imposing restrictions on cycle length has consequences as some words are not able to produce 'long' cycles. For example, the English noun *abacus* would produce no candidates if it were required to have 6-node cycles but correctly produces a new target if we allow for 5-node cycles:
Assuming that polysemic words trigger big contexts, we may argue that "small contexts" do not involve (many) polysemic words. This allowed us being less restrictive and permit shorter cycles (5 nodes) when dealing with small contexts. Note that in any case, we are dealing with cycles involving 5 languages. Fig 4 describes different examples. *Doll*, *wrist* and *forest* have big contexts; with 6, 8 and 9 known targets and imply a rather big number of nodes and edges. For *forest*, cycle computation rejects 73 words. These include the *non-forest* meanings introduced by polysemic words in the context. *Alpha*, *abacus* and *action* have small contexts. They only have 3 and 5 known translations and imply a rather small number of nodes and edges.

	k.targets	nodes	edges	p.targets	refused
alpha	3	6	14	1	1
abacus	3	16	33	1	11
academy	5	14	40	2	6
doll	6	58	135	10	41
wrist	8	43	100	6	28
forest	9	88	161	6	73

Note that in the case of *alpha* the system finds a 5-node cycle and identifies a potential target (*alfa-pt*) but this is rejected as it has a low degree and the cycle a low density. *Abacus*, also involves a 5-node cycle and a low degree target. In this case however, the density score is 0.8. Finally, a*cademy*, manages to provide two new targets as in both cases, they got a high density score.

score		cycle / potential target /edges
0.6	fail	alfa-gl, **alfa-pt-(2),** alfa-es, alfa-ca, **alpha-en**
0.8	ok	abako-eo, ábaco-es, **abac-oc-(2)**, àbac-ca, **abacus-en**
0.8	ok	akademio-eo, academia-es, **academia-oc-(2)**, acadèmia-ca, **academy-en**

0.9	ok	akademio-eo, academia-es, **académie-fr-(3)**, acadèmia-ca**, academy-en**

Table 14: *Small contexts* examples

Taking the *forest* example we discussed in section 5, we observe that when running the experiment with this new restrictive criteria the system still manages to produce 4 candidates (*bosque-pt, bois-fr, bòsc-oc and fort-fr*) and rejects *selva-es*:

score	cycle / potential target /edges
0.73	bosque-gl, **bosque-pt-(3)**, bosque-es, bosc-ca, arbaro-eo, **forest-en**
0.73	fraga-gl, bosque-es, bosc-ca, **bois-fr-(3),** arbaro-eo, **forest-en**
0.53	bosc-ca, arbaro-eo, fort-fr, **selva-es-(2)**, baso-eu, **forest-en**
0.73	fraga-gl, bosque-es, **bòsc-oc-(3)**, bosc-ca, arbaro-eo, **forest-en**
0.73	fraga-gl, bosque-es, **fort-fr-(3)**, bosc-ca, arbaro-eo, **forest-en**

Table 15 The *forest* example

To evaluate the productivity of the system when applying the new restrictive algorithm, we run again the experiment against the 5476 English nouns that provide some cycle in the EN-ES experiment and got the results in Table 16. Figures in italics are for the rejected targets (30%) and the T.d. column shows the number of edges for the involved target. Note that about 21% of the words did not produced any accepted cycle (those with score 0.00) either because the cycles had less than 5 nodes or because the target two edges. Few candidates (8%) were rejected because of the low degree of source/target and low density. Finally, 4579 candidates (70%) are accepted.

targets	%	score	T.d.	*targets*	*%*	*score*	*T.d.*
8	0.1%	0.70	2	*1354*	*21%*	*0.00*	*2*
214	3.3%	0.75	2	*133*	*2%*	*0.00*	*3*
50	0.8%	0.50	3	*25*	*0%*	*0.50*	*2*
39	0.6%	0.53	3	*220*	*3%*	*0.55*	*2*
253	3.9%	0.57	3	*20*	*0%*	*0.57*	*2*
107	1.6%	0.60	3	*27*	*0%*	*0.60*	*2*
173	2.6%	0.63	3	*213*	*3%*	*0.65*	*2*
657	10.0%	0.65	3	***1992***	*30%*		
13	0.2%	0.67	3				
926	14.1%	0.75	3				
687	10.5%	0.85	3				
87	1.3%	0.57	4				
685	10.4%	0.63	4				
680	10.3%	0.70	4				
4579	69.7%						

7. Discussion

Cycle computation produces a list of pair candidates (in any language) with a confidence score. The sample examples reported in section 5 demonstrate that cycle computation is quite productive both in terms of new translation candidates and in terms of new target languages involved.

In the experiments reported in section 6, candidates were validated against three 'reference' dictionaries. For the set of candidate pairs that could be checked in the reference dictionaries, validation demonstrates cycle computation correctly identifies them

In some cases, pair candidates were not found in the reference dictionaries, so no validation was possible for them. In the EN-ES experiment we argued that since the 'extra' candidates in the COSD case, were included in the original Apertium data, they were correct. In the EN-FR experiment we initiated a manual checking that, initially, led us to assume that 'extra' candidates were also correct. However, when checking the *doll/wrist* case (as representative examples of correlated polysemy, the worst scenario) we found that in such extreme cases, the system did not perform well and wrongly produced incorrect targets. This led us to be more restrictive when admitting cycles and producing new targets (essentially, we require longer cycles and impose some restrictions on node's degree).

The eventual experiment, performs well for the extreme case of correlated polysemy but, obviously, is less productive. Note, however that in the EN-ES experiment, cycle computation would automatically produce 4579 EN-ES new translation pairs which constitutes the **35.95% of the original EN-ES data.** Though productivity depends on source data, we understand that some adjustments can be applied to increase productivity for the 24% of words that fail to produce an 'accepted' cycle.

8. Conclusions

The experiments presented here exploit the properties of the Apertium RDF Graph, principally nodes' degree and cycle density, to automatically generate new translation relations between words, and therefore enrich existing Apertium bilingual dictionaries with new entries. We understand that successive executions would provide even better results in terms of 'coverage'. The results we got are still preliminary but promising and lead us to address the possibility to use cycles and nodes degree to identify potential oddities in the source data. If cycle density proves efficient when considering potential targets, we can assume that in dense graphs nodes with low degree may indicate a potential error.

Crucial in this such a scenario is the notion of **context of W** which allows focusing the computation on a limited sub-graph. Further refinements can be applied when setting the **context of W** not only the +/- language repetition mode we already applied but also limiting the context within a specific subset of languages.

9. Source data

Source data can be found at https://github.com/martavillegas/ApertiumRDF

10. Acknowledgements

This work has been partially supported by the LIDER FP7 European project (ref. 610782) and by the Spanish Ministry of Economy and Competitiveness through the project 4V (TIN2013-46238-C4-2-R), the

Excellence Network ReTeLe (TIN2015-68955-REDT) and the Juan de la Cierva program.

Part of the work was carried out within the FP7-ICT-2013-10 STREP project EUMSSI under grant agreement nº 611057, receiving funding from the European Union's Seventh Framework Programme managed by the REA (Research Executive Agency).

11. Bibliographical References

T. Flati and R. Navigli. "The CQC algorithm: Cycling in graphs to semantically enrich and enhance a bilingual dictionary." Proceedings of the Twenty-Third international joint conference on Artificial Intelligence. AAAI Press, 2013.

M. Forcada, M. Ginestí-Rosell, J. Nordfalk, J. O'Regan, S. Ortiz-Rojas, J. Pérez-Ortiz, F. Sánchez-Martínez, G. Ramírez-Sánchez, and F. Tyers. Apertium: a free/opensource platform for rule-based machine translation. Machine Translation, 25(2):127–144, 2011

J. Gracia, E. Montiel-Ponsoda, D. Vila-Suero, and G. Aguado-de-Cea. Enabling language resources to expose translations as linked data on the web. In Proc. of 9th Language Resources and Evaluation Conference (LREC'14), Reykjavik (Iceland), pages 409–413. ELRA, May 2014.

J. Gracia, M. Villegas, A. Gómez-Pérez, N. Bel. The Apertium Bilingual Dictionaries on the Web of Data. Under review in http://www.semantic-web-journal.net/content/apertium-bilingual-dictionaries-web-data, 2015

L. T. Lim, B. Ranaivo-Malançon, and E. K. Tang. Low cost construction of a multilingual lexicon from bilingual lists. Polibits, 43:45–51, 2011.

J. McCrae, G. Aguado-de-Cea, P. Buitelaar, P. Cimiano, T. Declerck, A. Gómez-Pérez, J. Gracia, L. Hollink, E. Montiel-Ponsoda, D. Spohr, and T. Wunner. Interchanging lexical resources on the semantic web. Language Resources and Evaluation, 46(4):701–719, 2012

S. Soderland et al. Compiling a massive, multilingual dictionary via probabilistic inference. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1. Association for Computational Linguistics, 2009.

K. Tanaka and K. Umemura. Construction of a bilingual dictionary intermediated by a third language. In COLING, pages 297–303, 1994