Coupling for Coreference Resolution in a Never-ending Learning System

F. Quécole[1], M.C. Duarte[2], E.R. Hruschka[1]

[1] Universidade Federal de São Carlos, Brazil
felipequecole@gmail.com, estevam@dc.ufscar.br
[2] Bradesco Bank S.A.
maisa.cristina.duarte@univ-st-etienne.fr

Abstract. The Never-Ending Language Learning (NELL) is a system that attempts learning to learn from the Web every day, in an autonomous way. Maintaining high precision is the key to keeping the NELL's learning active and improving day-by-day. One of the challenges for NELL system is to properly identify different noun phrases that denote the same concept in order to maintain the cohesion of the knowledge base. This article investigates the coupling as an approach for improving coreference resolution on NELL. For that, several coupled algorithms, and simple ensemble methods, considering semantic and morphologic features were compared with results previously obtained with no use of coupling. The results presented in this article confirm empirically that coupling strategy is a useful and good approach to achieve better coverage and accuracy in NELL's knowledge base.

Categories and Subject Descriptors: I.2.6 [**Artificial Intelligence**]: Learning

Keywords: coreference, coupling, machine learning, never-ending learning, ensemble

1. INTRODUCTION

This article is part of the studies for the Never-Ending Language Learner, also known as NELL [Mitchell et al. 2015] [1]. NELL is a never-ending learning system based on semi-supervised learning, which learns from extracting facts on the world wide web. NELL has been running since january 12th, 2010 and is located at Carnegie Mellon University (Pittsburgh, USA). NELL's goal is to learn new facts, improve its own knowledge base, 24 hours per day, 7 days a week, forever. NELL's knowledge base is available at `http://rtw.ml.cmu.edu/rtw/kbbrowser`. Keeping this context in mind, this article's main goal is to deal with coreference resolution on NELL. Thus, coreferences are different named entities which denote the same semantic meaning, eg: "Shaquille O'Neal" and "Shaq". This problem is common in information extraction systems, such as NELL [Carlson et al. 2010], and *TextRunner* [Yates et al. 2007], which both fail to identify coreference entities. Such a problem prevents the system from interpreting different entities with the same meaning. With this, the knowledge base can store entities such as: "Paris" and "City of Lights" as different instances, thus reducing the accuracy of the knowledge base and preventing a greater quantity of facts from being extracted as if using these entities as coreference. Coreferences, in NELL's context, can be calculated from named entities' morphologic and/or semantic features.

Morphology, in linguistics' context, as well as in this article's, studies word format in isolation, that is, disregarding the context in which the words are located. Semantics, in the context of this

[1] `http://rtw.ml.cmu.edu/rtw`

Thanks UFSCar (Federal University of São Carlos) for the infrastructure provided and Bradesco Bank S.A. for the financial support.

article, refers to the relations in the knowledge base in which the named entities take part. Currently, NELL's coreference solver is ConceptResolver [Krishnamurthy and Mitchell 2011], which considers only relations to calculate coreference. Take Table I as an example.

Table I. Relations' instances as semantic features

Relation / NE1	Shaquille O'Neal	Shaq O'Neal	Shaq	Stephen Curry
athleteAlsoKnownAs	Shaq O'Neal	Shaq	Shaq O'Neal	-
athleteLedSportsTeam	Suns	Suns	Suns	Warriors
athletePlaysForTeam	Suns	Suns	Suns	Warriors
athletePlaysInLeague	NBA	NBA	NBA	NBA
athletePlaysSport	Basketball	Basketball	Basketball	Basketball

Table I presents the possible instances for the named entities (NE1) (column), given the relations (rows). For instance, given the relation *athletePlaysSport* and the named entity (NE1) "Shaquille O'Neal", the result is "Basketball". The "-" indicates that that named entity does not take part in that relation. Named entities, in NELL's context, are instances of people, cities, objects, events, etc.

This article is organized as follows. Section 2 presents the main correlated works on the context of this article; in section 3, the coreference resolution through morphological features, semantic features and both (with or without coupling) is investigated; section 4 presents the results and discussion; section 5 presents the conclusions and future work

2. NEVER-ENDING LANGUAGE LEARNING

NELL is a system that aims to learn uninterruptedly from the web. The objective is to make NELL able to learn more and better every day (or iteration). At each iteration, the system promotes some of the new learned facts (the ones with higher confidence) and uses them as labeled data for the next iteration.

In order for learning to take place, NELL has as input an ontology, which is organized into categories, relations and seeds (examples) of both. The categories are the types of knowledge to be learned, such as: city(x), country(y), state(w), person(z), etc., while relations are relationships between categories, such as cityLocatedInCountry(city(x), country(y)), personBornInLocation(person(z),city(x)), cityCapitalOfCountry(city(x),country(y)), etc. For categories, for example, the seeds could have as values of x: São Carlos, Ribeirão Preto, Paris, etc. while for relations they could have as seeds of x and y, respectively: São Carlos & Brazil, Ribeirão Preto & Brazil, Paris & France, etc., and so on for the other categories and relations. Figure 1 illustrates a possible subset from the ontology.

NELL is based on the semi-supervised learning paradigm [Zhu 2010], thus being susceptible to semantic-shift [Curran et al. 2007], that is, as new facts are learned and incorporated into the knowledge base (and thus used as labeled data for other iterations), small semantic deviations can occur that if neglected, can accumulate and impair the system's learning. Such deviations could be caused by errors or concept changing over time (eg: the president of a country). In order to minimize the semantic-shift, NELL uses several coupled components as presented in [Carlson et al. 2009], which means that each component performs learning from a different view, and at the end, both results are combined for a more accurate decision.

In a few words, as presented in several publications about NELL [Mitchell et al. 2015; Duarte et al. 2016], coupling is an important key to keeping a never-ending system learning.

Some of NELL's major components are:

— CML (*Coupled Morphologic Learner*) [Carlson et al. 2009] identifies named entities through morphological analysis;

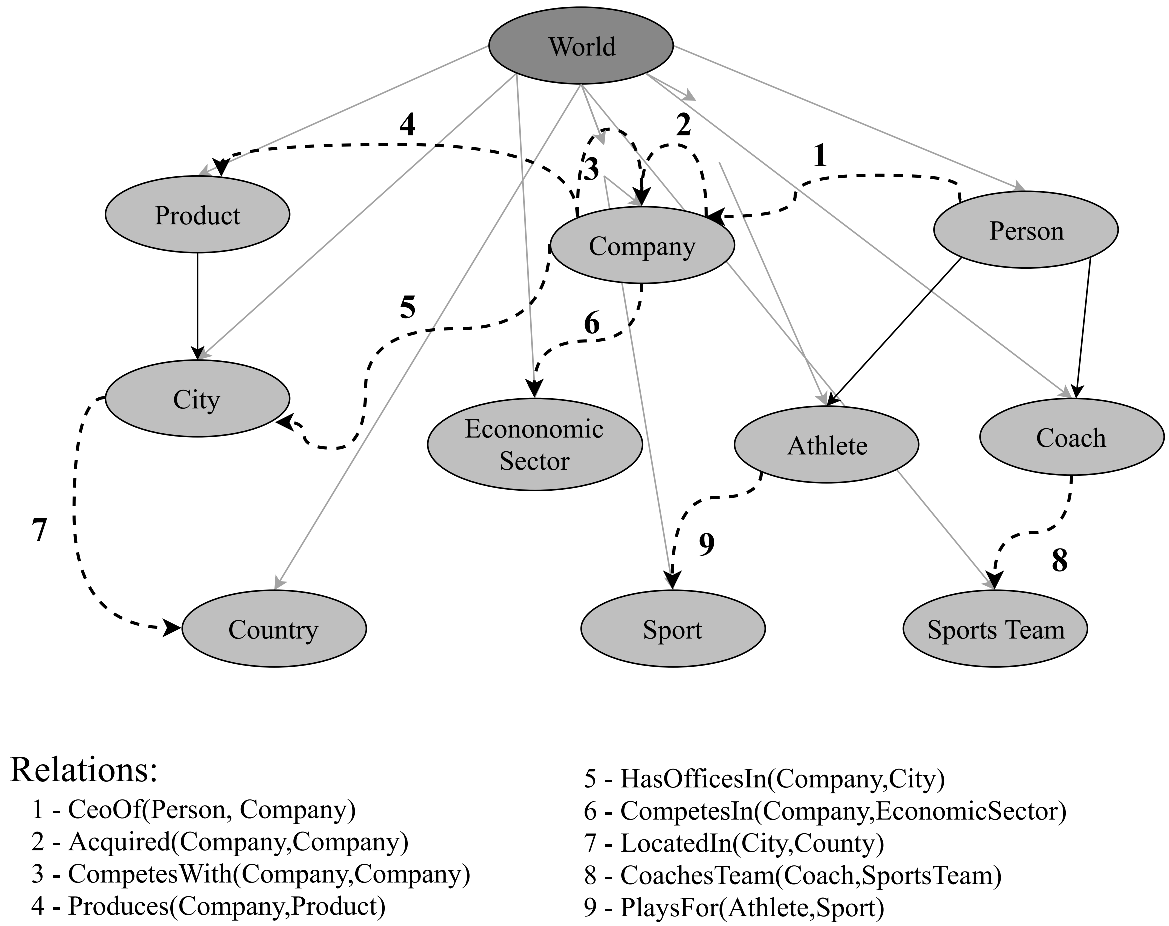

Fig. 1. Subset from NELL's ontology
[Duarte and Hruschka 2014]

— CPL (*Coupled Patterns Learner*) [Carlson et al. 2010] extracts named entities using textual patterns (and textual patterns using named entities as well);

— SEAL (*Coupled Set Expander for Any Language*) [Carlson et al. 2009] works similar to CPL, but using HTML patterns;

— *ConceptResolver* [Krishnamurthy and Mitchell 2011] NELL's current coreference solver. ConceptResolver, the main component studied in this article, performs clustering based on the similarities of semantic features, and thus decides if a pair of named entities are a coreference.

— OpenEval [Samadi et al. 2013] automatically evaluates the correctness of a predicate instance using the Web.

— PRA (*Path Ranking Algorithm*) [Mitchell et al. 2015] Infers new beliefs from the current knowledge base

— NEIL (*Never Ending Image Learner*) [Chen et al. 2013] learns from images related to the noun phrase

— OntExt (*Ontology Extender*) [Mitchell et al. 2015] considers every pair of categories in NELL's current ontology, to search for evidence of a frequently discussed relation between members of the category pair

NELL's architecture is presented at Figure 2:

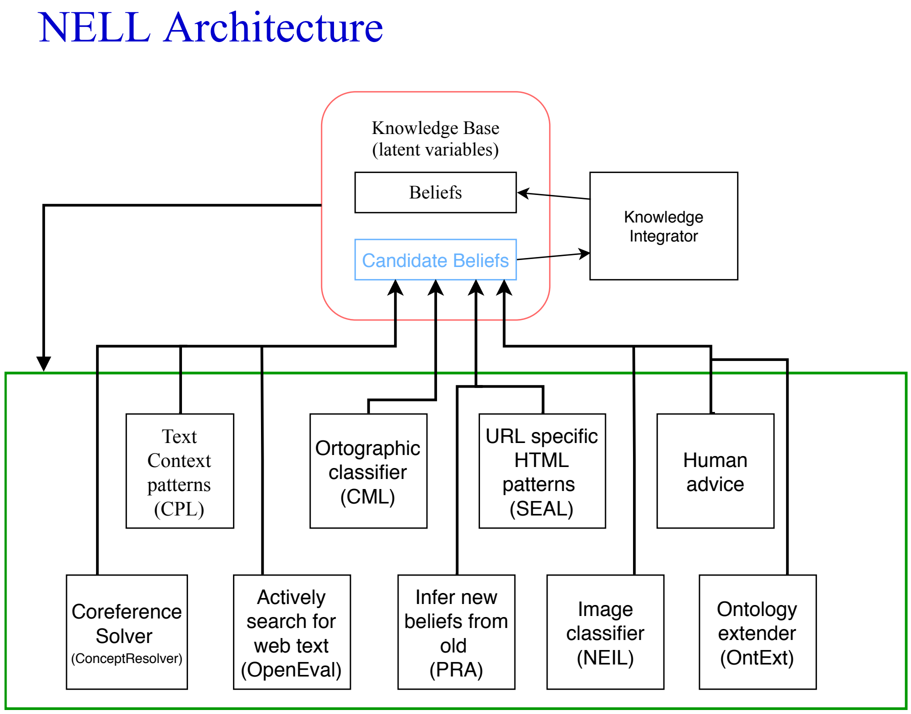

Fig. 2. NELL's architecture
Adapted from [Mitchell et al. 2015]

3. COREFERENCE RESOLUTION

NELL's current coreference component, the ConceptResolver, acts by grouping named entities that are candidates for coreference according to their occurrences only from the relations in the knowledge base. Therefore, ConceptResolver can find named entities, obtaining good results, based on semantic features, but it has difficulties with relations with missing values, as shown in the Table II.

Table II. Relations instances with missing values

Relation / NE1	**Shaquille O'Neal**	**Messi**	**Cristiano Ronaldo**	**Stephen Curry**
athleteHomeStadium	Talking Stick Resort Arena	-	Santiago Bernabeu	-
athletePlaysForTeam	-	Barcelona	Real Madrid	Warriors
belongsTo	-	Argentina	Portugual	-
athletePlaysInLeague	NBA	La Liga	-	NBA
athletePlaysSport	Basketball	Soccer	Soccer	Basketball
hasSpouse	-	-	Irina Shayk	-

Handling coreferences in never-ending learning systems, such as NELL, is important, since the negligence of these occurrences could cause a representation problem in the extracted knowledge. For example: the system would identify that "Shaquille O'Neal" is an athlete, as well as "Shaq" is an athlete, that both play basketball, and led Phoenix Suns' team, but the system would not know that both reference the same real-life entity. Neglecting this equivalence may cause the system to learn, for example, that Shaquille O'Neal played in the NBA, but not learn that Shaq also played, even though both terms refer to the same real-life entity. Which means that, even though the information is held by the system, it is not well-referenced, causing problems that could be solved by handling coreferences.

[Duarte and Hruschka 2014] presented a discussion on semantic features (based on Kownledge Base's (KB) relations) used by ConceptResolver and how the results could be improved with the addition of morphological features (based on KB's categories). Also in [Duarte and Hruschka 2014] the authors concluded that both approaches, semantic and morphological, when used together can, in fact, provide better results, positively impacting the accuracy and NELL's coverage. Furthermore, it was also pointed out that the independence between both approaches should be investigated as one more form of improvement of coreference resolution. For this, the coupling of both characteristics was pointed out as a possible solution, which is the approach investigated in this article, using ensemble classifiers.

An ensemble classifier is a set of base classifiers working together, and can be seen as a basic coupling method. For this article, the ensemble works in a way known as voting classifier. All the base classifiers do their jobs, and "vote" for one class. The class with most votes for that data is chosen. Another approach for the ensemble classifier used in this article was the one known as "Stacking" [Wolpert 1992], in which the base classifiers generate a "new dataset" used by a learning method (usually another classifier) to make the decision.

The KB incompleteness in relation instances can cause problems for the current approach used by NELL, that only considers relations as feature for coreference resolution. Therefore, the use of morphological features contribute in this matter, minimizing the problem.

In [Duarte and Hruschka 2014] the following morphological features were proposed:

(1) **String similarity:** real-valued contained in the interval [0,1]. The closer to 1, the greater the similarity between named entity 1 (ne1) and named entity 2 (ne2);
(2) **Different words:** binary feature, if value = 1, ne1 and ne2 have different quantity of words, otherwise, the value is set to 0;
(3) **Subset:** binary feature, it is set to 1 if ne1 is part of ne2, or if ne2 is part of ne1, otherwise, it is set to 0. Example: Subset(Shaquille O'Neal, Shaq) = 1, e Subset(Shaq, Curry)=0;
(4) **Acronym:** binary feature, it is set to 1 if one of the parameters is acronym for the other. Otherwise, it is set to 0. Example: Acronym(SP, Sao Paulo)=1, e Acronym(RJ, Shaquille O'Neal)=0;
(5) **Convington's String Similarity:** real-valued, contained in the interval [0,1], feature obtained applying Covington's algorithm [Covington 1996];
(6) **Proximity:** real-valued, contained in the interval [0,1], features obtained applying the Jaro-Winkler based algorithm. [Carpenter 2007].

In addition, the following attributes were used, which simulate the processing used by ConceptResolver:

(7) **Number of relations sharing the same instance value (SharingRelations):** integer value resulting from the sum of relations contained in the knowledge base that share the same value for ne1 and ne2. Take table I as an example: SharingRelations(Shaquille O'Neal, Shaq) = 5, and SharingRelations(Shaq, Stephen Curry) = 4;
(8) **Ratio of relations sharing the same instance value (SharingRelationsRatio):** real value contained in the interval [0,1] refers to the proportion between the number of relatios sharing the same value for ne1 and ne2, and the total number of relations in which both ne1 and ne2 take part. Taking table I as an example: SharingRelationsRatio(Shaquille O'Neal, Shaq) = 5/5 = 1, e SharingRelationsRatio(Shaq, Stephen Curry) = 2/4 = 0.5.

For this article, all 8 mentioned features were reproduced and recalculated. The morphological features were adapted from the related work [Duarte and Hruschka 2014]. The semantic features were adapted from the simulation of ConceptResolver [Krishnamurthy and Mitchell 2011], which is NELL's current coreference solver. Once all the features were selected, the next step was to create a dataset to test them. For this, an extraction method was implemented, which searches for coreference candidates in some specific relationships. The output of this method went through a manual selection, composing the dataset for conducting the experiments (dataset for training and testing)

In this article, the following relations (and its instances) were selected for composing the dataset: personAlsoKnownAs, athleteAlsoKnownAs, cityAlsoKnownAs. From which, 200 pairs of named entities, 100 true positives and 100 false positives, were selected and manually labeled. This amount corresponds to about 80% of these relations' instances (according to a query held on March 20th). It's possible to browse NELL's KnowledgeBase on `http://rtw.ml.cmu.edu/rtw/kbbrowser/`. After obtaining the features to be investigated and the dataset for training and test, some classifiers were selected and used, in order to evaluate how significant the selected characteristics are. The Python module Scikit-learn [Pedregosa et al. 2011] was used and the selected classifiers are listed further in this paper.

In order to investigate the improvement of results with the approach that explores the two views, the experiments contained in [Duarte and Hruschka 2014] were reproduced, using the dataset cited above, obtained for this article, which are:

— **Experiment 1 (exp1):** Considers both semantic and morphological features, that is, considers all 8 features mentioned above.
— **Experiment 2 (exp2):** Considers only morphological features, that is, only the first 6 mentioned above.
— **Experiment 3 (exp3):** Considers only semantic feature, that is, considers only the last 2 features mentioned above. This simulates ConceptResolver.

In order to verify improvement through an F-Measure analysis, a coupling of morphological and semantic characteristics was implemented through an ensemble [Dietterich 2002]. That is, a composition of independent classifiers that are improved through a voting process (Voting classifier). The ensemble presented in this work was implemented using the Python Sci-Kit [Pedregosa et al. 2011] module and consists of the grouping of four classifiers, applied separately for morphological and semantic data, resulting in a total of nine classifiers. The chosen classifiers were:

(1) Gaussian Naive-Bayes (GNB);
(2) K-Nearest Neighbors (KNN);
(3) Decision Tree;
(4) Random Forest;
(5) Logistic Regression
(6) Support Vector Classification
(7) Multilayer Perceptron;
(8) Gaussian Process Classifier;
(9) Ada Boost.

In order to investigate the results of the ensemble classifier, more experiments were performed:

— **Experiment 4 (exp4):** *ensemble* with the four of the classifiers with better scores in previous experiments (exp 1, 2 and 3) as its base classifiers, that is, KNN, Random Forest, AdaBoost and Decision Tree, experiment 4 uses the "Voting" approach;
— **Experiment 5 (exp5):** *ensemble* with the same classifiers as experiment 4 as base classifiers, but using the "Stacking" method [Wolpert 1992], with Bernoulli Naive-Bayes responsible for the final decision

— **Experiment 6 (exp6):** *ensemble* using the same method as experiment 5, but with Multilayer Perceptron Classifier responsible for the final decision

4. RESULTS AND DISCUSSION

The results obtained reinforce the conclusions discussed in [Duarte and Hruschka 2014] and, in addition, presents the behavior of the coupling approach using an ensemble classifier, which obtained better results than those presented by the cited authors.

Table III presents the results for each classifier, using cross-validation with 10 folds, taking into account experiments 1, 2 and 3. The emboldened values indicate which of the experiments performed best in terms of F-Measure (also known as F-Score).

Table III. Results for experiments 1, 2 and 3

<table>
<tr><th rowspan="3">Algorithm</th><th rowspan="3">Morphological</th><th rowspan="3">Semantic</th><th colspan="4">Confusion Matrix</th><th rowspan="3">F-Measure</th></tr>
<tr><th colspan="2">YES</th><th colspan="2">NO</th></tr>
<tr><th>Correct</th><th>Wrong</th><th>Correct</th><th>Wrong</th></tr>
<tr><td rowspan="3">Gaussian Naive Bayes</td><td colspan="2">exp1</td><td>59</td><td>41</td><td>97</td><td>3</td><td>0.707</td></tr>
<tr><td>exp2</td><td></td><td>55</td><td>45</td><td>99</td><td>1</td><td>0.680</td></tr>
<tr><td></td><td>exp3</td><td>91</td><td>9</td><td>19</td><td>81</td><td>0.670</td></tr>
<tr><td rowspan="3">KNN</td><td colspan="2">exp1</td><td>85</td><td>15</td><td>81</td><td>19</td><td>0.835</td></tr>
<tr><td>exp2</td><td></td><td>76</td><td>24</td><td>91</td><td>9</td><td>0.817</td></tr>
<tr><td></td><td>exp3</td><td>78</td><td>22</td><td>48</td><td>52</td><td>0.665</td></tr>
<tr><td rowspan="3">Decision Tree</td><td colspan="2">exp1</td><td>79</td><td>21</td><td>82</td><td>18</td><td>0.821</td></tr>
<tr><td>exp2</td><td></td><td>74</td><td>26</td><td>87</td><td>13</td><td>0.807</td></tr>
<tr><td></td><td>exp3</td><td>67</td><td>33</td><td>73</td><td>27</td><td>0.709</td></tr>
<tr><td rowspan="3">Random Forest</td><td colspan="2">exp1</td><td>84</td><td>16</td><td>83</td><td>17</td><td>0.831</td></tr>
<tr><td>exp2</td><td></td><td>72</td><td>28</td><td>83</td><td>17</td><td>0.789</td></tr>
<tr><td></td><td>exp3</td><td>67</td><td>33</td><td>71</td><td>29</td><td>0.695</td></tr>
<tr><td rowspan="3">Logistic Regression</td><td colspan="2">exp1</td><td>75</td><td>25</td><td>85</td><td>15</td><td>0.775</td></tr>
<tr><td>exp2</td><td></td><td>68</td><td>32</td><td>90</td><td>10</td><td>0.753</td></tr>
<tr><td></td><td>exp3</td><td>58</td><td>42</td><td>66</td><td>34</td><td>0.580</td></tr>
<tr><td rowspan="3">Support Vector Classification</td><td colspan="2">exp1</td><td>71</td><td>29</td><td>93</td><td>7</td><td>0.783</td></tr>
<tr><td>exp2</td><td></td><td>65</td><td>35</td><td>96</td><td>4</td><td>0.755</td></tr>
<tr><td></td><td>exp3</td><td>81</td><td>19</td><td>50</td><td>50</td><td>0.496</td></tr>
<tr><td rowspan="3">Multilayer Perceptron Classifier</td><td colspan="2">exp1</td><td>83</td><td>18</td><td>92</td><td>17</td><td>0.832</td></tr>
<tr><td>exp2</td><td></td><td>75</td><td>9</td><td>91</td><td>25</td><td>0.810</td></tr>
<tr><td></td><td>exp3</td><td>59</td><td>34</td><td>66</td><td>41</td><td>0.591</td></tr>
<tr><td rowspan="3">Gaussian Process Classifier</td><td colspan="2">exp1</td><td>79</td><td>15</td><td>85</td><td>21</td><td>0.810</td></tr>
<tr><td>exp2</td><td></td><td>71</td><td>8</td><td>92</td><td>29</td><td>0.783</td></tr>
<tr><td></td><td>exp3</td><td>56</td><td>35</td><td>65</td><td>44</td><td>0.516</td></tr>
<tr><td rowspan="3">Ada Boost Classifier</td><td colspan="2">exp1</td><td>84</td><td>17</td><td>83</td><td>16</td><td>0.837</td></tr>
<tr><td>exp2</td><td></td><td>77</td><td>18</td><td>82</td><td>23</td><td>0.789</td></tr>
<tr><td></td><td>exp3</td><td>67</td><td>27</td><td>73</td><td>33</td><td>0.674</td></tr>
</table>

In Table III, referring to the reproduction of the experiments presented in [Duarte and Hruschka 2014], but using the new base extracted for this article, in all classifiers better results were obtained with the joint use of morphological and semantic features (exp1). Classifiers that used only semantic features (exp3) had the worst results, causing many false positives to appear, while "morphological-only" classifiers (exp2) tend to result in fewer false-positives and more false-negatives. Such behavior reinforces the idea that the two approaches have independent errors

Differently from the results presented in [Duarte and Hruschka 2014], where only semantic classifiers presented better results in detriment of only morphological classifiers, in this article the results were the opposite. This difference is attributed to the fact that both experiments were performed on small, but significant, datasets. As already mentioned, this dataset corresponds to about 80% of the entities present in relations that indicate coreference. In this article, the extraction was done automatically through a Java implementation of a tool to accomplish the task, which makes easier for future experiments and investigations, unlike the [Duarte and Hruschka 2014] approach, in which the authors manually extracted the base.

Aiming to investigate the coupling approach with the new dataset extracted from NELL and after the reproduction of the experiments presented in [Duarte and Hruschka 2014], three experiments (Exp4, Exp5 and Exp6) were carried out using ensemble classifiers. The results of such an approach are presented in Table IV, which presents the confusion matrix for the three experiments, as well their F-Measure.

Table IV. Ensemble experiments results

	YES		**NO**		**F-Measure**
	Correct	**Wrong**	**Correct**	**Wrong**	
Exp4	86	13	87	14	0.866
Exp5	84	9	91	16	0.871
Exp6	89	15	85	11	0.867

Analyzing the results obtained in Table IV, one can conclude that the coupled approach obtains more accurate results than the semantic and morphological approach experienced as a single set of attributes. For this case, McNemar's significance test [Dietterich 1998] was run, getting a p-value below our threshold of 5% (p-value=0.021) when comparing Experiments 1 (using Random Forest Classifier) and 4. Table V demonstrates the contingency table used for McNemar's algorithm. It's also possible to conclude that the "Stacking" method obtained slightly better results in terms of F-measure, but when checking for statistical significance a p-value of 0.15 was found, considering Experiments 4 and 5, failing to prove significance.

Table V. Contingency table for McNemar's test

	Exp1 Correct	**Exp1 Incorrect**
Exp4 Correct	164	9
Exp4 Incorrect	1	26

In order to check the consistence of experiments' results, all the experiments were performed 50 times (always using cross-validation with 10 folds) and the metrics were calculated along with their standard deviation (SD), the results are shown on Table VI

Table VI. Experiments 4, 5 and 6 results

	F-Measure		**Accuracy**	
	Mean	**SD**	**Mean**	**SD**
Exp4	0.870	0.011	0.869	0.013
Exp5	0.872	0.011	0.871	0.013
Exp6	0.863	0.010	0.864	0.011

From Table VI it's possible to conclude that, even though experiment 5's performance was slightly better, both approaches (Voting or Stacking) achieved good results, slightly better than just having all the features (morphological and semantic) but without coupling.

All the following studies and experiments performed in this article were performed using the Voting method, because it was easier to gather and visualize the results individually, and, as shown above, the final results were sufficiently similar to the other approaches.

Table VII presents some instances evaluated by the ensemble classifier, in which the numbers refer to the classifiers presented in section 3, T and F represent the votes of each classifier: coreference and non-coreference, respectively; blue means that the classification is correct, and red and emboldened, incorrect.

Table VII. Instances evaluated by the ensemble classifier

NE1	NE2	Morphological				Semantic				Ensemble Class
		1	2	3	4	1	2	3	4	
Canadian Jewish Congress	**CJC**	T	T	T	T	F	F	F	F	Coreference
New York City	**Big Apple**	F	F	F	F	T	T	T	T	Coreference
Ho Chi Minh	**Saigon**	F	F	F	F	F	T	T	T	**No-coreference**
Madrid	**Rio**	F	T	T	T	F	F	F	F	No-coreference
David Lee	**David Robinson**	T	T	T	T	F	F	F	F	**Coreference**

Table VII highlights the collaboration of each morphological and semantic classifier for the resolution of the transferences. In the instance that relates "New York City" and "Big Apple" it can be noted that only morphology would not be able to identify the relationship between terms, demonstrating that semantic attributes are still the most important in this calculation. But when those fail, as in the case of the "Canadian Jewish Congress" and "CJC", it is interesting to consider the morphological attributes in the resolution

The errors presented in this approach are due to the fact that ensemble is implemented by a process called "simple voting", that is, the results of all classifiers have the same weight, which allows the propagation of some errors. As in the case of "David Lee" and "David Robinson", which are terms morphologically similar but unrelated. Although all semantic classifiers showed that there would be no relation between terms, the instance was erroneously classified with reference, since all morphological classifiers indicated coreference.

Another case derived from the above problem involves " Ho Chi Minh " and " Saigon ", the terms are not morphologically close, but the semantic classifiers pointed to a possible relationship, except for one of them. As a result, the instance was also incorrectly classified, this time as no-coreference.

Table VIII presents a small sample of evaluated instances and their classifications according to each method, being:

— **M**: Morphological-only method (exp2);
— **S**: Semantic-only method (exp3);
— **MS**: Mophological + Semantic method (exp1);
— **CMS**: Coupled morphological + semantic method(exp5).

Table VIII. Instances examples for each approach

EN1	EN2	Class	M	S	MS	CMS
Sin City	**Vegas**	YES	NO	YES	NO	YES
Los Angeles	**LA**	YES	YES	NO	NO	**YES**
Calcutta	**Kolkata**	YES	YES	NO	YES	**YES**
Santiago	**Campo Grande**	NO	NO	YES	NO	NO
Los Banos	**LA**	NO	YES	NO	YES	**NO**

The results presented in Table VIII prove that the coupled method was able to achieve better results. It was also demonstrated that the importance of considering morphological and semantic attributes in order to improve the results based on independent errors. Therefore, the approach proposed in this article has achieved good results and could be used as a basis for future studies on NELL's coreference analysis through coupling methods.

5. CONCLUSION AND FUTURE WORK

From the experiments described in this article, it is evident that the coupled use of morphological and semantic characteristics presents better results in the analysis of coreferences, considering that the two approaches present independent errors. The methods consider different aspects of each named entity and therefore have independent errors. For example, if only morphological aspects are taken into account, one can evaluate "compliment" and "complement" as coreference, although they do not denote the same meaning. Or, considering "King of Pop"and "Michael Jackson"as not being coreferences. But when considering also semantic aspects, the relation between the terms would be clear. This demonstrates the importance of independent errors by providing one approach to "correct" the other's weaknesses, causing an improvement in terms of accuracy.

The use of coupling solutions for some other problems in NELL's context has already been presented in [Carlson et al. 2009], [Hruschka Jr et al. 2013] and [Krishnamurthy and Mitchell 2011]. Likewise in this current article, better results were also achieved, thus proving that coupling is a good ally, also, for the resolution of coreferences. In this article, the coupled use of morphological and semantic features presented better results than those obtained by the ConceptResolver [Krishnamurthy and Mitchell 2011], the current NELL component. With this, this article presents and proposes a new method for solving NELL's coreferences. The proposed method has not yet been implemented in the system, but it is one of the future work approaches.

Finally, as a continuation of this work, it is proposed to investigate and implement a new method for coupled classification based on base classifiers' independent results and errors.

REFERENCES

Carlson, A., Betteridge, J., Hruschka, Jr., E. R., and Mitchell, T. M. Coupling semi-supervised learning of categories and relations. In *Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing*. SemiSupLearn '09. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1–9, 2009.

Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka, Jr., E. R., and Mitchell, T. M. Toward an architecture for never-ending language learning. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*. AAAI'10. AAAI Press, pp. 1306–1313, 2010.

Carpenter, B. Lingpipe for 99.99% recall of gene mentions. In *Proceedings of the Second BioCreative Challenge Evaluation Workshop*. Vol. 23. -, pp. 307–309, 2007.

Chen, X., Shrivastava, A., and Gupta, A. Neil: Extracting visual knowledge from web data. In *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, pp. 1409–1416, 2013.

Covington, M. A. An algorithm to align words for historical comparison. *Computational linguistics* 22 (4): 481–496, 1996.

Curran, J. R., Murphy, T., and Scholz, B. Minimising semantic drift with mutual exclusion bootstrapping. In *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*. Vol. 6. pp. 172–180, 2007.

Dietterich, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation* 10 (7): 1895–1923, 1998.

Dietterich, T. G. Ensemble learning. *The handbook of brain theory and neural networks* vol. 2, pp. 110–125, 2002.

Duarte, M. C. et al. Leitura da web em português em ambiente de aprendizado sem-fim, 2016.

Duarte, M. C. and Hruschka, E. R. Exploring two views of coreference resolution in a never-ending learning system. In *Hybrid Intelligent Systems (HIS), 2014 14th International Conference on*. IEEE, pp. 273–278, 2014.

Hruschka Jr, E. R., Duarte, M. C., and Nicoletti, M. C. Coupling as strategy for reducing concept-drift in never-ending learning environments. *Fundamenta Informaticae* 124 (1-2): 47–61, 2013.

Krishnamurthy, J. and Mitchell, T. M. Which noun phrases denote which concepts? In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*. HLT '11. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 570–580, 2011.

Mitchell, T., Cohen, W., Hruschka, E., Talukdar, P., Betteridge, J., Carlson, A., Dalvi, B., Gardner, M., Kisiel, B., Krishnamurthy, J., Lao, N., Mazaitis, K., Mohamed, T., Nakashole, N., Platanios, E., Ritter, A., Samadi, M., Settles, B., Wang, R., Wijaya, D., Gupta, A., Chen, X., Saparov, A., Greaves, M., and Welling, J. Never-ending learning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI'15. AAAI Press, pp. 2302–2310, 2015.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* 12 (Oct): 2825–2830, 2011.

Samadi, M., Veloso, M. M., and Blum, M. Openeval: Web information query evaluation., 2013.

Wolpert, D. H. Stacked generalization. *Neural networks* 5 (2): 241–259, 1992.

Yates, A., Cafarella, M., Banko, M., Etzioni, O., Broadhead, M., and Soderland, S. Textrunner: Open information extraction on the web. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*. NAACL-Demonstrations '07. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 25–26, 2007.

Zhu, X. pp. 892–897. In C. Sammut and G. I. Webb (Eds.), *Semi-Supervised Learning*. Springer US, Boston, MA, pp. 892–897, 2010.