Supervised Machine Learning for Predicting the Meaning of Verb-Noun Combinations in Spanish

Olga Kolesnikova and Alexander Gelbukh

Center for Computing Research, National Polytechnic Institute
Mexico City, 07738, Mexico
kolesolga@gmail.com, www.gelbukh.com

Abstract. The meaning of such verb-noun combinations as *take care*, *undertake work*, *pay attention* can be generalized as DO what is designated by the noun. Likewise, the meaning of *make a decision*, *provide support*, *write a letter* can be generalized as MAKE what is designated by the noun. These generalizations represent the meaning of certain groups of verb-noun combinations. We use supervised machine learning algorithms to predict the meanings DO, MAKE, BEGIN, and CONTINUE of previously unseen verb-noun pairs. We evaluate the performance of the applied algorithms on a training set using 10-fold cross-validation technique. The learnt models have also been evaluated on an independent test set and the predictions have been checked manually to determine the accuracy of the classifiers. The obtained results show that supervised machine learning methods achieve significant accuracy and can be used for semantic annotation of verb-noun combinations.

Keywords: lexical functions, verb-noun combinations, meaning representation by means of hypernyms, supervised machine learning.

1 Introduction

The meaning of individual words can be described by definitions in conventional dictionaries for human usage like Longman Dictionary of Contemporary English or the Merriam-Webster English Dictionary. Often, most frequent words have many senses. For example, Longman Dictionary of Contemporary English [5] gives 47 senses for the verb *take*, 44 for *make*, the number of senses for *have* reaches 49, but *play* looks very poor with only 10 senses! Combinations of verbs with prepositions, called phrasal verbs, like *take after*, *make over*, *have on* etc. are not counted as separate senses otherwise the number of senses would have grown tremendously!

Taking a careful look at definitions of the previously given verbs, one can notice that these verbs have some meanings in common. Note, that we have used word definitions from the Longman Dictionary of Contemporary English mentioned above; therefore in this Section, when referring to the dictionary we mean the Longman Dictionary of Contemporary English.

Coming back to the fact of meaning repetitions in verb definitions, we now give a few examples of verbs which have the meaning DO STH (STH = something) among other senses. First, let us consider the verb *take*. The dictionary gives the following definition of *take* in the sense DO STH: ‘a word meaning to do something used with many different nouns to form a phrase that means: “do the actions connected with the nouns”: take a walk / take a bath / take a breath / take a vacation.’ The second example is the verb *make*. In the dictionary, it also has the sense DO STH followed by the comment: ‘used with some nouns to mean that someone performs the action of the noun: make a decision / mistake.’ Thirdly, even the verb *have* which is typically used in the sense *possess*, can acquire the meaning DO STH in combination with some nouns. In this meaning, *have* is described as ‘a word meaning to do something, used in certain phrases: have a look / walk / sleep / talk / thing / a holiday / bath / shower’.

Lastly, let us consider the verb *play*. One of its meanings given in the dictionary is ‘to take part in a game or sport’ like golf, chess, etc. Though the exact phrase DO STH or the exact word DO is not encountered in the definition of *play*, we look for the definition of *to take part* in the dictionary and find: *to take part* is ‘to do an activity, sport etc. with other people’. Therefore it can be affirmed that *play* also has DO as one of its senses, because in the definition of *play*, we can substitute *to take part* by ‘to do an activity, sport etc. with other people’.

We will call the meaning DO STH, or just DO, the generalized meaning of the verbs *take*, *make*, *have*, and *play*, since DO is used in the first, more general, part of the verb definitions. Table 1 gives other examples of the generalized meaning DO. For clarity and illustration, verbs are given in combination with nouns.

Table 1. Verbs with the meaning DO.

Verb	Dictionary definition of the sense generalized as DO
give somebody / sth a smile / laugh / shout / push	do something – to smile, laugh, shout etc.: *He gave me a quick smile and a hug*. | *Ooh, the baby just gave a kick*!
conduct a survey / experiment / inquiry etc	to carry out a particular process, especially in order to get information or prove facts: *The company conducted a survey to find out local reaction to the leisure center*.
carry sth ***out***	to do something that needs to be organized and planned: *They are carrying out urgent repairs*. | *A survey is now being carried out nationwide*. | *It won’t be an easy plan to carry out*.
ask (a question)	to say or write something in order to get an answer, a solution, or information: *That kid’s always asking awkward questions*.
teach	to give lessons in a school, college, or university: *The guy’s been teaching in France for 3 years now*.

Likewise, other generalized meaning can be determined. In this work, we are interested in generalized meanings DO, MAKE, BEGIN and CONTINUE. We do not give formal definitions for MAKE, BEGIN and CONTINUE but illustrate them with examples in Tables 2, 3, and 4, respectively. The examples are chosen to represent the given meanings in an exact and comprehensive way.

It is a generally accepted fact that the meaning of an individual word depends on its context, i.e. the words it is used with in corpora. This fact is also true in the case of generalized meanings that we have selected. Verbs acquire these meanings when collocate with nouns belonging to a particular semantic group, for example, a group denoting actions. If verb-noun combinations are annotated with the meanings DO, MAKE, BEGIN or CONTINUE, this annotation disambiguate both the verb and the noun. Word sense disambiguation is one of the most important and challenging tasks of natural language processing, and therefore semantic annotation of verb-noun combinations is a task of significant relevance.

It should be noted here that the concept of generalized meaning we propose here is close to the notion of lexical functions developed by the Meaning-Text Theory. Lexical function is a mapping from one word (called **keyword**, for example, *decision*) to another it collocates with in corpora (called **lexical function value**). This mapping is further characterized by the meaning of semantically homogenuous groups of values and by typical syntactic patterns in which lexical function values are used with their respective keywords in texts. For the keyword *decision*, the lexical function Oper_1, meaning 'do, perform, carry out', gives the value *make*. That is, to express the meaning 'do, or perform, a decision', one says in English *make a decision*. The formalism of lexical functions is intended to represent fixed word combinations, or collocations like *make a decision*, *give a lecture*, *lend support*, etc. For more information on lexical functions consult [6, 7]. We do not apply the formalism of lexical functions as it is. Our purpose is to predict semantic contents of verb-noun combinations, and the meanings we have chosen, are not exactly the meanings of lexical functions though have some resemblance to them. Another difference is that lexical functions describe collocations, but generalized meanings are present in collocations as well as in free word combinations. Section 3 gives details concerning state-of-the-art research on lexical functions.

The rest of the paper is organized as follows. Section 2 formulates our task. Section 3 gives a summary of related work. Section 4 explains what data was used in the experiments. Section 5 describes methodology. We present results and discuss them in Section 6. Section 7 outlines conclusions and future work.

Table 2. Verbs with the meaning MAKE.

Verb	Dictionary definition of the sense generalized as MAKE
create	to make something exist that did not exist before: *Her behaviour was creating a lot of problems*.
cause	to make something happen: *Heavy traffic is causing long delays on the freeway*.
build	to make something, especially a building or something large: *Are they going to build on this land?*
write	to produce a new book, poem, song etc.
produce	to make things to be sold: *Gas can be produced from coal*.

Table 3. Verbs with the meaning BEGIN.

Verb	Dictionary definition of the sense generalized as BEGIN
start	to begin doing something: *start learning German / work*
enter	to start working in a particular profession or organization: *Andrea is studying law as a preparation for entering politics*.
introduce	be the start of; if an event introduces a particular period or change, it is the beginning of it: *The death of Pericles in 429 BC introduced a darker period in Athenian history*.
launch	to start something, especially an official, public, or military activity that has been carefully planned: *launch a campaign / appeal / inquiry*
become	to begin to be something: *He became King at the age of 17.*

Table 4. Verbs with the meaning CONTINUE.

Verb	Dictionary definition of the sense generalized as CONTINUE
keep	to continue to have something and not lose it or get rid of it: *No, we're going to keep the house in Vermont and rent it out.*
maintain	to make something continue in the same way or at the same high standard as before: *Britain wants to maintain its position as a world power.*
pursue	to continue doing an activity or trying to achieve something over a long period of time: *Kristin pursued her acting career with great determination.*
sustain	to mak something continue to exist over a period of time: *The teacher tried hard to sustain the children's interest.*
run	to continue to be officially able to be used for a particular period of time: *The contract runs for a year.*

2 Task

The task of our work is to examine performance of supervised learning methods for prediction of the meanings DO, MAKE, BEGIN, and CONTINUE in Spanish verb-noun combinations. We train classifiers on a manually compiled corpus of verb-noun pairs annotated with the above given meanings. After building classification models on the training data, the models are tested for prediction of the meanings on unseen data. The data used for testing the models are of two types. The first type of the testing data is a part of the training set which is divided into the training section and the test section applying the 10-fold cross-validation technique. The second type of testing data is an independent test set build on a corpus other than the corpus used to construct the training set. The details concerning data can be found in Section 4.

3 Related Work

The meaning of word combinations is often represented as a semantic relation between individual words that constitute word combinations. We will give a short review of two lines of research devoted to semantic relations.

The first line is work on automatic detection of lexical functions mentioned in the Introduction. Lexical functions are semantic relations which hold between constituents of fixed word combinations, or collocations. Collocations may have different syntactic structures, and the verb-noun pattern is one of these structures. L.Wanner [16, 17] made experiments to classify Spanish verb-noun pairs according to nine lexical functions with the meaning 'perform, experience, carry out something, 'cause the existence of something, 'begin to perform something, 'continue to perform something', etc. Verb-noun pairs were divided in two groups. In the first group, nouns belonged to the semantic field of emotions; in the second groups nouns were field-independent. For classification, the following supervised learning algorithms were applied: Nearest Neighbor technique, Naïve Bayesian network, Tree-Augmented Network Classification technique and a decision tree classification technique based on the ID3-algorithm. As a source of information for building the training and test sets, hyperonymy hierarchy of the Spanish part of EuroWordNet [15, 12] was used. The average f-measure of about 70% was achieved in these experiments. The best results for field-independent nouns were shown by ID3 algorithm (f-measure of 0.76) for the lexical function with the meaning 'cause (by the noun functioning in utterances as the verb's direct object) something to be experienced / carried out / performed' and by the Nearest Neighbor technique (f-measure of 0.74) for the lexical function with the meaning 'perform / experience / carry out something'.

The second line of research on semantic relations in word combinations deals with automatic assignment of semantic relations to English noun-modifier pairs in [8, 9]. Though in our work, verb-noun combinations are treated, we believe that the principles of choosing data representation and machine learning techniques for detection of semantic relations between a noun and a modifier can also be are used to detect semantic relations in verb-noun pairs. The underlying idea is the same: learning the meaning of word combinations. In [8, 9], the researchers examined the following relations: causal, temporal, spatial, conjunctive, participant, and quality. They used two different data representations: the first is based on WordNet relations, the second, on contextual information extracted from corpora. They applied memory-based learning, decision tree induction and Support Vector Machine. The highest f-score of 0.847 was achieved by C5.0 decision tree to detect temporal relation based on WordNet representation.

4 Data

Verb-noun pairs were extracted automatically from the Spanish Web Corpus [11] by the Sketch Engine [4] and ranked by frequency. Thus we obtained a list of 83, 982 pairs. From this list, we have taken the first one thousand pairs and processed them manually as follows.

First, we removed all fallacious combinations extracted from the Spanish Web Corpus automatically due to parsing errors. Erroneous pairs included, for instance, past participles or infinitives instead of nouns, or contained symbols like --, « , © instead of words. The total number of erroneous pairs was 61, so after their removal the list contained 939 pairs.

Secondly, we disambiguated each verb and noun, annotating them with word senses of the Spanish WordNet [15, 12]. For some verb-noun pairs, relevant senses were not found in the above mentioned dictionary, and the number of such pairs was 39. For example, in the combination *dar cuenta*, 'give account', the noun *cuenta* means *razón*, *satisfacción de algo*, 'reason, or satisfaction of something'. This sense of *cuenta* is taken from *Diccionario de la Lengua Española*, 'Dictionary of the Spanish Language' [2]. Unfortunately, this sense is absent in the Spanish WordNet so the expression *dar cuenta* was left without sense annotation. All combinations that could be not annotated with senses of the Spanish WordNet were removed from the list.

After the first two steps, 900 verb-noun pairs were left in the list. We have looked through the list and annotated all relevant combinations with the meanings DO, MAKE, BEGIN, and CONTINUE. We found 280 pairs with the meaning DO, 112 pairs with the meaning MAKE, BEGIN was encountered in 25 pairs, and CONTINUE was observed to be the most rare meaning with only 16 verb-noun pairs. Thus the total number of verb-noun pairs annotated with four meanings was 433, and 467 pairs had meanings other than DO, MAKE, BEGIN, CONTINUE. All 900 pairs were included in the training sets. Table 5 demonstrates examples of the data. The examples are given as they are encountered in the list built automatically, so the nouns are used without articles or quantifiers.

We build four training sets, one for each of the four meanings. All training sets included the same 900 examples which were marked differently depending on the meaning chosen for a given set. For example, the training set for DO included 280 positive instances marked as the class "yes" and the rest of the examples (620 instances) were marked as the class "no", i.e. these were instances of the meanings MAKE, BEGIN, and CONTINUE, as well as the verb-noun pairs with meaning other that DO, MAKE, BEGIN, CONTINUE.

Table 5. Examples of verb-noun pairs.

Meaning	**Examples**	
	Spanish	**English lit. translation**
DO	*hacer justicia* *realizar actividad* *dar beso*	*do justice* *realize activity* *give kiss*
MAKE	*hacer ruido* *establecer criterio* *encontrar solución*	*make noise* *establish criterion* *find solution*
BEGIN	*iniciar proceso* *tomar iniciativa* *adoptar actitud*	*initialize process* *take initiative* *adopt attitude*
CONTINUE	*mantener control* *llevar vida* *seguir curso*	*maintain control* *lead life* *follow course*

Lastly, for each verb and noun in the training sets, we extracted all hyperonyms from the Spanish WordNet. We represented each verb-noun pair as a set of all hyperonyms of the noun and all hyperonyms of the verb. Both constituents of verb-noun pairs were considered as zero-level hyperonyms, they were also included in the set of hyperonyms.

To build an independent test set, we extracted 5181 verb-noun pairs from the Spanish Treebank Cast3LB [1], a corpus other than the corpus used to construct the training sets. To evaluate the performance of classifiers, we used the test set in the following ratios: 100%, 75%, 50%, and 25%.

We did not disambiguate verb-noun pairs for the test sets manually. Instead, for each verb-noun, we built all possible verb-noun combinations of all senses in the Spanish WordNet. As an example, let us consider the pair *representar papel*, 'represent role'. The verb *representar* has 12 senses in the Spanish WordNet, and the noun *papel*, 5. This gives totally 60 combinations of *representar* and *papel* (12 multiplied by 5). Remember, that the test data included totally 5,181 verb-noun pairs which resulted in 96,079 instances in the test set.

The training and test sets were formatted according to Attribute-Relation File Format (ARFF) [14] to be accessible by machine learning methods described in

Section 5. Every hyperonym was presented as an attribute with two possible values, “1” if a corresponding hyperonym is encountered in a particular verb-noun pair, and “0” if it is not. Thus each verb-noun pair was represented as a vector of zeros and ones. The last attribute was a categorical feature with two possible values, “yes” if a corresponding verb-noun pairs has the meaning that is to be learnt by classifiers, and “no” if it is not.

5 Methodology

Our approach is based on supervised machine learning algorithms as implemented in the WEKA version 3-6-2 toolset [13, 3, 18]. We performed two groups of experiments. In the first group of experiments, we evaluated the prediction of the meanings DO, MAKE, BEGIN, and CONTINUE on the training sets using 10-fold cross-validation technique. In the second group of experiments, the same meanings were predicted for the instances of an independent test set. Table 6 lists all classifiers we experimented with.

Table 6. Classifiers.

Classifier	Classifier	Classifier
AODE	ClassificationViaClustering	VFI
AODEsr	ClassificationViaRegression	ConjunctiveRule
BayesianLogisticRegression	CVParameterSelection	DecisionTable
BayesNet	Dagging	JRip
HNB	Decorate	NNge
NaiveBayes	END	OneR
NaiveBayesSimple	EnsembleSelection	PART
NaiveBayesUpdateable	FilteredClassifier	Prism
WAODE	Grading	Ridor
LibSVM	LogitBoost	ZeroR
Logistic	MultiBoostAB	ADTree
RBFNetwork	MultiClassClassifier	BFTree
SimpleLogistic	MultiScheme	DecisionStump
SMO	OrdinalClassClassifier	FT
VotedPerceptron	RacedIncrementalLogitBoost	Id3
Winnow	RandomCommittee	J48
IB1	RandomSubSpace	J48graft
IBk	RotationForest	LADTree
KStar	Stacking	RandomForest
LWL	StackingC	RandomTree
AdaBoostM1	ThresholdSelector	REPTree
AttributeSelectedClassifier	Vote	SimpleCart
Bagging	HyperPipes	

6 Experimental Results

6.1. Experiments on the Training Sets

The purpose of our experiments was to evaluate performance of 68 classifiers on the training sets using 10-fold cross validation technique. The best five results for predicting the "yes" class are presented in Tables 7, 8 (remember, "yes" and "no" classes are explained in Section 4 alongside with other details about data). P stands for precision, R for recall and F for *f*-measure. Together with the best results, Table 7, 8 show performance of the four classifiers most frequently used in natural language processing, i.e. support vector machine (implemented in WEKA as SMO), C4.5 decision tree learner (J48 in WEKA), Naive Bayes algorithm, and nearest-neighbor instance-based learner (IB1 in WEKA). These 4 classifiers are left in the tables ranked by *f*-measure so it can be seen what algorithm is better for detecting each meaning.

It is a common practice that in classification experiments, the performance of rules.ZeroR classifier is considered as the baseline. ZeroR is a trivial algorithm that always predicts the majority class. But in our training sets the majority class is always the class of negative examples. Remember, that the overall number of positive and negative instances in the training sets is 900, though the largest number of positive instances is 280 for the meaning DO which still is much less then the number of negative instances (620 in the case of DO). Therefore, ZeroR does not classify any test instances as positives, which always gives recall of 0 and undefined precision. For this reason, ZeroR should not be considered as the baseline.

Table 7. Performance of WEKA Classifiers on DO and MAKE training sets.

DO				MAKE			
Classifier	P	R	F	Classifier	P	R	F
PART	0.898	0.857	**0.877**	JRip	0.726	0.706	0.716
SimpleCart	0.901	0.853	0.876	SimpleCart	0.728	0.688	0.708
BLR	0.875	0.872	0.874	LADTree	0.706	0.706	0.706
Bagging	0.884	0.857	0.870	REPTree	0.721	0.688	0.704
BFTree	0.903	0.838	0.869	BFTree	0.730	0.670	0.699
...	...	...	...	...	...	...	...
SMO	0.856	0.872	0.864	SMO	0.689	0.651	0.670
J48	0.876	0.850	0.863	J48	0.747	0.541	0.628
NaiveBayes	0.762	0.711	0.735	IB1	0.532	0.376	0.441
IB1	0.566	0.759	0.648	NaiveBayes	0.535	0.211	0.303

Table 8. Performance of WEKA Classifiers on BEGIN and CONTINUE training sets.

BEGIN				CONTINUE			
Classifier	P	R	F	Classifier	P	R	F
Prism	0.778	0.737	0.757	Ridor	0.813	0.813	0.813
FT	0.762	0.667	0.711	REPTree	0.857	0.750	0.800
SMO	0.824	0.583	0.683	LWL	0.857	0.750	0.800
VFI	0.750	0.625	0.682	EnsembleSelection	0.857	0.750	0.800
NNge	0.750	0.625	0.682	RandomSubSpace	0.857	0.750	0.800
...	...	...	...	...	...	...	...
JRip	0.667	0.500	0.571	J48	0.750	0.750	0.750
IB1	0.818	0.375	0.514	SMO	0.786	0.688	0.733
NaiveBayes	0.000	0.000	0.000	IB1	0.500	0.313	0.385
Prism	0.778	0.737	0.757	NaiveBayes	0.000	0.000	0.000

The best classifiers for prediction of the meaning DO is PART, for the meaning MAKE, JRip, for BEGIN, Prism, and for CONTINUE, Ridor. All four classifiers are rule based classification algorithms. Inductive rule learning use separate-and-conquer strategy. It means, that a rule that works for many instances in the class is identified first, then the instances covered by this rule are excluded from the training set and the learning continues on the rest of the instances. These learners are efficient on large, noisy datasets. Our training sets included 900 instances represented as vectors of the size 1109 attributes, and rule induction algorithms performed very well.

The best state-of-the-art result for predicting a lexical function with the meaning 'cause' is f-measure of 0.76 given by ID3 algorithm [17]. In our experiments, the best f-measure of 0.877 was shown by PART for the meaning MAKE. However, such a comparison is not fair, since our task was to predict the meanings DO, MAKE, BEGIN, CONTINUE but not lexical functions as explained in the Introduction.

6.2. Experiments on the Test Sets

Some of the best classifiers displayed in Tables 7, 8 were evaluated on an independent test set built as described in Section 4. Tables 9, 10 present the results for these classifiers. We listed the values of precision, recall and *f*-measure for each classifier in this way: <precision>|<recall>|<f-measure>; BLR in the column **Classifier** stands for BayesianLogisticRegression. As we explain below, the test sets had such a big size that some classifiers failed to make predictions within a reasonable time period. For such classifiers, we put N/A instead of metrics as for the other classifiers.

It was mentioned in Section 4, that since we did not disambiguate verb-noun pairs in the test sets, for each pair we build the number of instances equal to the number of senses for the verb multiplied by the number of senses for the noun. This has given us 96079 instances and 10544 attributes in 100% test set, 73021 instances and 9495 attributes in 75% test set, 48904 instances and 8032 attributes in 50% test set, and

22254 instances and 5857 attributes in 25% test set. SimpleCart, FT, LWL had difficulties in predicting the value of the class variable on test sets of sizes more than 25%. Among these three classifiers, SimpleCart was better because this algorithm was effective enough to process a 75% and 50% set. SimpleCart and FT are decision tree algorithms, and LWL is a nearest-neighbor instance-based learner. Note, that almost all the best classifiers that could process a full-size test set, belong to the class rules. BayesianLogisticRegression also performs well and the only algorithm of the class trees that did not experience time problems was LADTree.

Table 9. Performance of WEKA classifiers on the test set.

Meaning	**Classifier**	**Test set size**	
		100%	**75%**
DO	PART	0.261|0.864|0.400	0.304|0.864|0.382
	SimpleCart	N/A	N/A
	BLR	0.178|0.830|0.293	0.212|0.818|0.337
MAKE	JRip	0.231|0.662|0.342	0.189|0.662|0.294
	SimpleCart	N/A	0.168|0.662|0.268
	LADTree	0.285|0.676|0.401	0.168|0.676|0.269
BEGIN	FT	N/A	N/A
	SMO	0.331|0.793|0.467	0.567|0.793|0.661
	NNge	0.302|0.724|0.426	0.567|0.724|0.636
CONTINUE	Ridor	**0.799**|0.480|0.600	**0.993**|0.480|0.647
	REPTree	0.581|0.480|0.526	**0.820**|0.480|0.606
	LWL	N/A	N/A

Table 10. Performance of WEKA classifiers on the test set.

Meaning	**Classifier**	**Test set size**	
		50%	**25%**
DO	PART	0.245|0.864|0.382	0.162|0.852|0.272
	SimpleCart	0.405|0.864|0.551	0.281|0.852|0.423
	BLR	0.205|0.818|0.328	0.145|0.807|0.246
MAKE	JRip	0.189|0.662|0.294	0.174|0.662|0.276
	SimpleCart	0.203|0.662|0.311	0.177|0.662|0.279
	LADTree	0.144|0.676|0.237	0.177|0.676|0.281
BEGIN	FT	N/A	0.409|0.724|0.523
	SMO	0.464|0.793|0.585	0.404|0.793|0.535
	NNge	0.603|0.724|**0.658**	0.451|0.724|0.556
CONTINUE	Ridor	**0.958**|0.480|0.640	**1.000**|0.480|**0.649**
	REPTree	0.694|0.480|0.567	0.667|0.480|0.558
	LWL	N/A	**1.000**|0.480|**0.649**

As it is seen from Tables 9, 10, the best precision was shown by Ridor. This method (Ridor = RIpple-DOwn Rule learner) have been developed for knowledge acquisition where it is hard to add a new rule and be sure that it would not cause the inconsistency of the rules generated before. Ridor algorithm is different from covering algorithms for constructing the rule set; instead it generates exceptions for the existing rules that work within the confines of these rules thus not affecting other rules. Then it iterates on the exceptions for the best solution. This scheme allowed the classifier to reach 100% precision. Unfortunately, it can not boast the best recall which is only 0.649 for the meaning CONTINUE on a 25% test set. Still, it is the second best recall in our experiments on test sets. The top recall is 0.658 shown by NNge for the meaning BEGIN on a 50% test set.

Another classifier that gives the best precision of 100% is LWL when performing predictions for the meaning CONTINUE on a 25% test set. But, like Ridor, it shows the same low recall of 0.658. However, a high precision of Ridor and LWL makes them appropriate for fulfilling the tasks where precision is of special importance, for example, for automatic construction of dictionaries.

7 Conclusions and Future Work

We have shown that it is feasible to apply machine learning methods as implemented in the WEKA toolkit for predicting the meaning of unseen Spanish verb-noun collocations. In particular, we trained classifiers to assign the meanings DO, MAKE, BEGIN and CONTINUE to a previously unseen verb-noun pair.

Verb-noun pairs were represented as sets of hyperonyms for both the verb and the noun. As our experiments have shown, hyperonyms function sufficiently well as features distinguishing between the meanings we chosen to be predicted by classifiers.

The best f-measure achieved in our experiments is 0.877 using the training set and 10-fold cross-validation technique. This is significantly higher than the previously reported result of 0.740 for f-measure, though the comparison is not fair because we looked for the meaning which is similar to the meaning predicted in [17], but not the same one. The highest f-measure achieved in the experiments on an independent test set was only 0.658. This could be explained by the fact that the best ratio between the training set and the test set has not yet been found by us. More experiments on test sets of various sizes are needed.

In the future, we plan to test other classification methods that were not examined in our experiments as well as to work with data extracted from a raw corpus and lemmatized [10]. We also plan to study the effect of other features, such as WordNet glosses and to make experiments with word space models representing various similarity measures between word combinations. We will experiment with different ratios between the training set and the test set.

Acknowledgements. We are grateful to Adam Kilgarriff and Vojtěch Kovář for providing us a list of most frequent verb-noun pairs from the Spanish Web Corpus of the Sketch Engine, www.sketchengine.co.uk.

The work was done under partial support of Mexican Government: SNI, COFAA-IPN, PIFI-IPN, CONACYT grant 50206-H, and SIP-IPN grant 20100773.

References

1. Civit, M., Martí, M.A.: Building Cast3LB: A Spanish Treebank. In: Research on Language and Computation, vol. 2(4), pp. 549-574. Springer, Netherlands (2004)
2. Diccionario de la Lengua Española. Real Academia Española, Madrid (2001)
3. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten I. H.: The WEKA Data Mining Software: An Update. SIGKDD Explorations, Volume 11, Issue 1 (2009)
4. Kilgarriff, A., Rychly, P., Smrz, P., Tugwell, D.: The Sketch Engine. In Proceedings of EURALEX 2004, pp. 105–116 (2004)
5. Longman Dictionary of Contemporary English. Third Edition. Longman Group Ltd, Essex, England (1995)
6. Mel'čuk, I. A.: A Theory of the Meaning-Text Type Linguistic Models (in Russian). Nauka Publishers, Moscow (1974)
7. Mel'čuk, I. A.: Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In: Wanner, L. (ed) Lexical Functions in Lexicography and Natural Language Processing, pp. 37–102. Benjamins Academic Publishers, Amsterdam, Philadelphia, PA (1996)
8. Nastase, V., Szpakowicz S.: Exploring noun-modifier semantic relations. In: 5th International Workshop on Computational Semantics (IWCS-5), pp. 285–301. Tilburg, Netherlands (2003)
9. Nastase, V., Sayyad-Shiarabad, J., Sokolova, M., Szpakowicz, S.: Learning noun-modifier semantic relations with corpus-based and wordnet-based features. In: Proceedings of the Twenty- First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference. AAAI Press (2006)
10. Sidorov, G.: Lemmatization in automatized system for compilation of personal style dictionaries of literature writers. In: Word of Dostoyevsky, pp. 266-300. Russian Academy of Sciences, Moscow (1996)
11. Spanish Web Corpus, http://trac.sketchengine.co.uk/wiki/Corpora/ SpanishWebCorpus/, last viewed June 02, 2010
12. Spanish WordNet, http://www.lsi.upc.edu/~nlp/web/index.php?Itemid=57&id=31&option=com_content&task=view, last viewed June 02, 2010
13. The University of Waikato Computer Science Department Machine Learning Group, WEKA download, http://www.cs.waikato.ac.nz/~ml/weka/index_downloading.html, last viewed June 02, 2010
14. The University of Waikato Computer Science Department Machine Learning Group, Attribute-Relation File Format, http://www.cs.waikato.ac.nz/~ml/weka/arff.html, last viewed June 02, 2010
15. Vossen P. (ed): EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers, Dordrecht (1998)
16. Wanner, L.: Towards automatic fine-grained classification of verb-noun collocations. Natural Language Engineering, vol. 10(2), pp. 95–143. Cambridge University Press, Cambridge (2004)
17. Wanner, L., Bohnet, B., Giereth, M.: What is beyond Collocations? Insights from Machine Learning Experiments. EURALEX (2006)
18. Witten, I. H., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2^{nd} Edition. Morgan Kaufmann, San Francisco (2005)