Procesamiento del Lenguaje Natural, Revista nº 61, septiembre de 2018, pp. 23-30 recibido 26-03-2018 revisado 02-05-2018 aceptado 03-05-2018

Detecting the Central Units of Brazilian Portuguese argumentative answer texts

Detección de las unidades centrales para textos de respuesta argumentativa en Portugués-Brasileño

Kepa Bengoetxea[1], Mikel Iruskieta[1], Juliano Antonio[2]
[1]University of the Basque Country (UPV/EHU). Ixa Group
[2]Universidade Estadual de Maringá (UEM). Funcpar.
kepa.bengoetxea@ehu.eus, mikel.iruskieta@ehu.eus, jdantonio@uem.br

Abstract: Understanding or writing properly the main idea or the Central Unit (CU) of a text is a very important task in exams. So, detecting automatically the CU may be of interest in language evaluation tasks. This paper presents a CU detector based on machine learning techniques for argumentative answer texts in Brazilian Portuguese. Results show that the detection of CUs following machine learning techniques in argumentative answer texts is better that those using rules.
Keywords: central unit, RST, argumentative answer texts

Resumen: Comprender o escribir correctamente la idea principal o Unidad Central (UC) de un texto es una tarea muy importante en los exámenes. Así, la detección automática de la UC puede ser de interés en las tareas de evaluación del lenguaje. Este artículo presenta un detector de UCs basado en aprendizaje automático para textos de respuesta argumentativa en Brasileño. Los resultados muestran que la detección de las UCs utilizando aprendizaje automático en brasileño y textos de respuesta argumentativa obtienen mejores resultados que los basados en reglas.
Palabras clave: unidad central, RST, textos de respuesta argumentativa

1 Introduction

Information about discourse structure can improve the realization of complex linguistic tasks such as automatic summarization, text generation, segmentation, information extraction, sentence compression, automatic translation, paraphrasing and even evaluating texts. To do such tasks, we need annotated corpora to develop and test an automatic discourse parser. Regarding Brazilian Portuguese, CST News Corpus (Aleixo and Pardo, 2008) is a corpora with information about discourse structure, which consists of news texts annotated with Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) and Dizer (Pardo and Nunes, 2004) is an automatic discourse analyzer for this language.

Both the corpora and the parser are based on RST, a theory that investigates the coherence relations which arise from the combination between text spans (Mann and Thompson, 1988). A very important notion for the theory is nuclearity. In asymmetric relations, the nuclear span is the member of the pair that is more essential to the writer's purpose than the others. As RST relations can be recursive, nuclearity works at different levels (top and bottom) of the relational discourse structure, which is hierarchical due to these asymmetric relations.

RST diagrams are represented as trees (henceforth RS-trees), and the central unit (henceforth CU; the most salient node of the rhetorical structure tree, which is at the top of the RS-tree) is an elementary discourse unit (henceforth EDU) which is not satellite of any other unit or text span.

The corpus studied was created with texts written by candidates for university entrance exams. The texts of the corpus were produced as an answer to the question "What's the secret of Vestibular: intelligence, effort or luck?". According to Menegassi (2011), argumentative answer genre belongs to scholar/academic sphere. It is initiated by the resumption of the question followed by the answer to the question, which is the thesis defended by the author. The remainder of the text presents arguments that support the thesis in order to try to convince or persuade

ISSN 1135-5948. DOI 10.26342/2018-61-2

the reader.

The aim of this paper is to build up a CU detector based on machine learning techniques for argumentative answer text in Brazilian Portuguese. The identification of the CU is relevant for the training of the Portuguese teachers committee which corrects the texts manually or also to the students who need some help in indicating the central idea in this genre, which can be crucial in their future,[1] and it is also crucial to develop a better discourse parser.

The option for building a corpus and not using an annotated existing corpus is derived from the goal of developing an application for the automatic detection of the CU of argumentative answer texts and also to show that an ADA have to detect first the CU and after elaborate the RS-tree.

2 Theoretical background and related work

In Antonio and Santos (2014) description of the rhetorical structure of argumentative answer genre, the initial statement is the CU of the text and its development is the satellite. In the corpus investigated by the authors, EVIDENCE relation is very common to be held between the satellite and the nucleus as the writer intends to increase the reader's belief on the content of the nucleus. ELABORATION relation also occurs in the corpus when writers intend to present additional information to the content of the nucleus. It must be noticed that not only these relations may be held between the central unit and the remainder of the text in argumentative answer texts. An example of argumentative answer text of our corpus is presented in Figure 1.

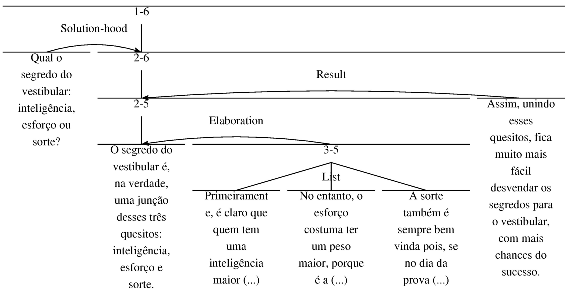

Figure 1: RS-tree of an argumentative answer text [M21294]

[1]The gold standard data can be consulted at http://ixa2.si.ehu.eus/rst/pt/.

There, after the question that the student has to answer ($span_1$), the text is divided into 5 spans. $Span_2$ is the CU, i.e., the unit which presents the main idea of the text. As the other spans are satellites regarding the CU, the arrows point towards $span_2$. $Spans_{3-5}$ hold ELABORATION relation with the nucleus (which is also the CU or the RS-tree). In ELABORATION relation, the satellite provides additional details about the elements of the nucleus (Mann and Thompson, 1988). In other words, the writer provides more information about the three secrets to achieve success in the entrance exams presented in $span_2$. In $span_3$ he elaborates intelligence, in $span_4$ he elaborates effort and in $span_5$ he elaborates luck. The relation held among $span_3$, $span_4$ and $span_5$ is LIST, a multinuclear rhetorical relation. Finally, $span_6$ is a RESULT satellite. The question that the student has to answer is related to the argumentative answer as a SOLUTION-HOOD relation, which is at the top of the RS-tree.

Regarding the identification of the most important discourse unit within the framework of RST, some applications have been developed: Pardo, Rino and Nunes (2003) developed an extractive summarizer for texts of any domain and Pardo an Nunes (2004) created DiZer, an ADA, both of them for Brazilian Portuguese.

The automatic discourse analyzer DiZer for BP (adapted especially to scientific texts) found a CU for this example[2] shown in Example (1):

(1) *Qual o segredo do vestibular: inteligência, esforço ou sorte?*

DiZer detected the question itself of the argumentative text as the CU, but if we introduce only the argumentative text in DiZer, the CU is the following: '*O segredo de o vestibular é, em a verdade, uma junção de esses três quesitos:*'.[3] The difference with the manual annotation is because of a different segmentation. DiZer left out the following part of text: *inteligência, esforço e sorte* from the $Span_1$ (in this case EDU_1). Whereas the CU for this text is: *O segredo do*

[2]There are several repositories and ways to use DiZer, we have used the *Portuguese by Dizer* and a greedy method to built the tree.

[3] English translation: The secret of vestibular is, in fact, a junction of three features:

vestibular é, na verdade, uma junção desses três quesitos: inteligência, esforço e sorte.[4]

Another similar program for detecting the most important idea in BP is GistSumm (Pardo, Rino, and Nunes, 2003). Example (2) was summarized with GistSumm at a 0.80 compresion rate.

(2) [*No entanto, o esforço costuma ter um peso maior, porque é a partir dele que os vestibulandos descobrem a melhor forma de estudar, muitas vezes abrem mão das horas de lazer para passarem mais tempo com os livros, enfim, quem é esforçado entende que não adianta ficar acomodado, é preciso dedicação para que se alcance o fim desejado.*]$_4$

GistSumm extracts the most important sentences using a ranking system based on words and sentence position. In the case of Example (2), GistSumm extracted the text span$_4$ in which the writer highlights the most important factor: 'effort'. Although the writer mentioned in the beginning of the text that the secret of *Vestibular* is a junction of the three factors (intelligence, effort and luck), in the span extracted by GistSumm, the writer assigns a bigger weight to effort.

There are similar works that detect the most important ideas of texts in different languages and domains: for English abstracts (Burstein and Marcu, 2003), for Spanish scientific texts (Bengoetxea and Iruskieta, 2018) and for Basque abstracts (Bengoetxea, Atutxa, and Iruskieta, 2017) developed a CU detector for scientific abstracts. For Brazilian Portuguese Iruskieta et al. (2016) used a rule-based automatic detector and the system gets a F-score 0.553 in the test dataset of this corpus.

3 Methodology

3.1 Corpus and annotators

The corpus used in this paper consists of 100 texts written by candidates for Summer 2013 entrance exams[5] at *Universidade Estadual de Maringá* (UEM). There are excerpts the candidates can base upon to write the texts demanded by the instructions. On Summer 2013 the question to be answered by candidates was metadiscoursive. They had to write about the factors that lead to success in Vestibular (university entrance exams). The instructions were: "*As a candidate, write, using up to 15 lines, an argumentative answer to the question '**What is the secret of Vestibular: intelligence, effort or luck?**'*. You can base upon the information of the excerpts, but you cannot copy them".

Corpus	Texts	Tokens	EDUs	CUs
Train	60	8,499	846	69
Test	40	6,511	576	50
Total	100	15,010	1,422	119

Table 1: Corpus information

The gold standard we created contains 1,422 EDUs and 100 texts, each with its CUs (see Table 3.1). The task's difficulty to find the CU has been calculated as follows: $Difficulty = \frac{CUs}{EDUs}$ where the nearer it is from 1 the easier it is to determine the CU. Test dataset is more difficult, because there are 11.77% more texts with multiple CU (in total 20%) and 5.2% less CUs in the EDU$_1$ position (in total 60%).

This corpus was divided into 2 non-overlapping datasets: the first 60 texts as a training dataset and the last 40 texts as test dataset.

The annotation phases were as follows:

i. A corpus of 100 argumentative answer texts was collected.

ii. Four annotators segmented the texts manually into EDUs and, afterwards, a super annotator harmonized the segmentation.

iii. Four annotators determined the CU of each text, and finally the texts were harmonized.

3.2 Linguistic Features (LF)

Following Antonio (2015), the development of the CU detector was based on the frequency of some lexical and grammatical items or indicators which were, therefore, chosen as features (see Table 2).

The first feature is a list of nouns with a high frequency in the CU. They are re-

[4]English translation: The secret of vestibular is, in fact, a junction of three features: intelligence, effort and luck.

[5]The exams are available at `http://www.vestibular.uem.br/2013-V/uemV2013p2g1.pdf`.

lated to the meaning of junction or combination, such as *junção* ‘junction’, *combinação* ‘combination’, *união* ‘union’, *conjuntura* ‘conjuncture’, *miscigenação* ‘miscigenation’ (which was misused by the writer), *mistura* ‘mixture’, *soma* ‘sum’ and *mescla* ‘mix’.

Regarding verbs, two types of verbs were also used to characterize the CUs: (*i*) copula verbs: *ser* ‘to be’ and *estar* ‘to be’, and (*ii*) evidential verbs which express propositional attitude (Dall’Aglio-Hattnher, 2007) such as *acreditar* ‘to believe’, *crer* ‘to believe’ and *pensar* ‘to think’ in first person singular.

Another group of features are bonus words: adverbs and adverbial phrases. These were also used by the writers in some CUs. It is important to remark that all epistemic adverbs used by writers were asseverative in an attempt to make their propositions more credible.

In argumentative answer genre, it is expected that writers resume the question before the answer. Thus, feature “title words” contains the words which were in the question that the writers were supposed to answer.

Finally, we found that the likelihood of a CU occurring at the beginning of the answer was quite high in the annotated data. To account for this, we used one feature that reflected the position of each EDU in argumentative answer texts.

Another group of features regards words which have low frequency in the CU and they are signals for other discourse structures (stop words). Thus, they can be used as cues in order not to identify the EDUs in which they appear as CU. It is plausible the fact that the conclusion may be stronger than the initial statement in terms of expressiveness. The conclusion is usually started by a discourse marker or a finisher expression such as *portanto* ‘therefore’, *enfim* ‘ultimately’, *a partir disso* ‘taking this into account’, *sendo assim* ‘thus’. Besides that, the fact that the answer and the stronger arguments are restated, and this makes the conclusion seem more assertive than the initial statement. The same case happens when the writer repeats the question in the text, this kind of segment can be detected with an question mark “?” or interrogative pronoun (in this case, *qual* ‘what’).

Table 2 summarizes all the features we used in machine learning techniques.

Group	Subgroup	Words
Nouns	junction of factors	junção, combinação, conjuntura, miscigenação, mistura, mix, soma, união, mescla
Verbs	copula	ser, estar
	evidential	acredito, creio, penso
Bonus words	epistemic	indubitavelmente, certamente, decerto, seguramente, obviamente, naturalmente, asseguradamente, inquestionavelmente, positivamente, decisivamente, incontestavelmente
	adverbial	sem dúvida, com certeza, de certeza, na verdade
Title words	resumption	segredo, vestibular
	factors	inteligência, esforço, sorte
Segment position		the position of each segment in the text
Not in CU	question	question mark(?) and qual
Non CU connectors	conclusion	portanto, enfim, sendo assim, por isso, contudo, pois, a partir de, de este maneira, me o qual, então, e que, de este modo, assim, porém

Table 2: Features extracted by a linguist from the training dataset

These features in argument answer texts are very different from other works that analyze other genre (Bengoetxea, Atutxa, and Iruskieta, 2017). So, we think that the indicators of the CUs are sensible to domain and genre. But there are some features that are common in most of them, such as: title words, segment position, epistemic adverbs, copula verbs and evidential verbs in first person singular.

3.3 Automatic Features (AF)

To detect the best features to tag the CU automatically we performed the following steps:

- We converted each segment words into a set of attributes representing word occurrence information and we created a set of 1000, 5000 and 15000 words (attributes) using the training data. We represented each segment by an array of words. Finally, the training set dictionary obtained using this scheme contains 1000 features.
- We converted all letters to lower case.
- We followed bag of words approach and used tokens (unigrams, bigrams, trigrams and fourgrams) as features, where a classification instance is a vector of tokens appearing in the segmented text.
- We added segment position and title word occurrence information to the feature vector. Using weka’s “string to word vector”, text was converted into feature vector using TF-IDF as feature.
- We removed noise feature. In general, the basic idea is to search through all

possible combinations of attributes in the data to find which subset of features works best for prediction. Removal is usually based on some statistical measures, such as segment frequency, information gain, Chi square or mutual information. We have tested the two most effective feature selection methods: *a*) Chi square and *b*) information gain using different set of attributes: 50, 100, 500 and 1000. Lastly, we performed all the classifiers using Chi square using a set of 100 attributes.

– Finally, the training set dictionary obtained using this scheme contains 100 features; the same dictionary was used for the test set.

4 *The systems*

In this paper, we conducted experiment in the Weka toolkit by using several supervised learning classifiers based on Support Vector Machines (SVM), Artificial Neural Networks, Bayes, Decision Tree and Rule-based method on the CU detector.

With the aim of selecting the best classifier, we have trained several learning classifiers on the indicators defined in Table 2 following Antonio (2015) using 10-fold cross-validation. The best systems are: *i*) Sequential Minimal Optimization (SMO) (Platt, 1998), *ii*) Multinomial Naive Bayes (MNB) and *iii*) Bernoulli Naive Bayes (BNB) system.

We have compared the results of three systems with a baseline and a rule-based automatic CU detector. A simple, but powerful baseline is based on the position of the given EDU into the whole document. The position is an important indicator, because we found that the likelihood of a CU occurring at the beginning of the answer was 65.2% in the training set. So we consider that the first segment is the only CU of the text as our baseline. The best rule-based system related in Iruskieta et al. (2016) also uses EDU position to detect the CU in argumentative answer texts. The best features in this system were: the EDU position (from 1 to 2) and at least 3 nouns or 3 title nouns of Table 2.

Our system has a module with 2 different stages to post-process the results from the classifiers:

i) The first stage deselects the conclusion segment (Post1): The most frequent error that the classifier performs in the cross-validation dataset is to select the conclusion segment as CU. So, if a EDU was considered as CU by the system and if it starts by a conclusion discourse marker (for example, ‘therefore’, ‘thus’, ‘ultimately’), then the CU is deselected.

ii) The second stage select at least one CU (Post2): Sometimes the systems classify all the segments of a text as non-CU or the first post-processing stage has deselected the CU. Depending on the classifier, we can apply 2 different techniques to select at least one CU:

  – A statistical post-processing to select the CU. In the case of MNB and BNB, the classifier always returns the probability of an EDU to be labelled as CU. So, the statistical post-processing use this value to select at least the most likely EDU to be labeled as CU.

  – The first EDU as CU. In the case of SMO, we consider that the first EDU is a unique CU of the text, but in the case that the first segment has an interrogative mark, the second EDU will be chosen by the system as the CU of the text. We also applied this technique in BNB and MNB systems, selecting the best one in each case.

Both stages (Post1 and Post2) are applied sequentially to process the outputs of the classifiers.

5 *Results*

We estimated the performance of ours systems using Linguistic Features (LF) and Automatic Features (AF). We partitioned the 60 texts into 10 groups. We trained 10 times on 9/10 of the labeled data and evaluated the performance on the other 1/10 of the data.

The evaluation results in Table 3 show the average performance of our classifier using traditional recall (Rec.), precision (Prec.), and F-score (F_1) metrics.

As we reported in Table 3.1, out of a total of 846 EDUs there are 67 CUs on the 10-fold cross-validation dataset (0.079 diffi-

System		D	C	E	M	P	R	F_1
B.line		C	45	15	24	0.750	0.652	0.697
		T	24	16	26	0.600	0.480	0.533
Rule-based		C				0.824	0.627	0.712
		T				0.778	0.429	0.553
LF+ BNB	W	C	48	21	21	0.695	0.695	0.695
		T	27	21	23	0.562	0.540	0.551
	P1	C	48	18	21	0.727	0.695	0.711
		T	27	13	23	0.675	0.540	0.600
	P2	C	48	19	21	0.716	0.695	0.705
		T	27	16	23	0.627	0.540	0.580
LF+ SMO	W	C	46	7	23	0.867	0.666	**0.754**
		T	25	3	25	0.892	0.500	**0.640**
	P1	C	46	6	23	0.884	0.666	**0.760**
		T	25	1	25	0.961	0.500	**0.657**
	P2	C	47	13	22	0.783	0.681	**0.728**
		T	29	11	21	0.725	0.580	**0.644**
AF+ SMO	W	C	42	12	27	0.777	0.608	0.682
		T	19	9	31	0.678	0.380	0.487
	P1	C	41	6	28	0.872	0.594	0.706
		T	19	4	31	0.826	0.380	0.520
	P2	C	46	15	23	0.754	0.666	0.707
		T	27	15	23	0.642	0.540	0.586
AF+ MNB	W	C	51	23	18	0.689	0.739	0.713
		T	25	21	25	0.543	0.500	0.520
	P1	C	49	15	20	0.765	0.710	0.736
		T	25	11	25	0.694	0.500	0.581
	P2	C	50	20	19	0.714	0.724	0.719
		T	28	17	22	0.622	0.560	0.589

Table 3: Results obtained on cross-validation (C) and test (T) sets without any post-process (W), with the first stage of the post-process (P1) and with the both stages of the post-process (P2)

culty)[6] and 576 EDUs there are 49 CUs on the test dataset (0.085 difficulty).

Table 3 shows results obtained on the 10-fold cross-validation and test sets using *i*) a baseline (all the first segments are considered as the unique CU of each text), *ii*) the best rule-based system (Iruskieta, Antonio, and Labaka, 2016) and *iii*) the best different machine learning methods using linguistics features and automatic features. These machine learning methods are MNB, BNB and SMO, each of them has three stages applied sequentially: *i*) the initial stage without any post-process (Without), *ii*) a first post-process stage (Post1) and *iii*) a second post-process stage (Post2). We can say that the rule-based detector has similar results in comparison with system's with Post1, because the system does not ensure that each text has a least one CU and all the conclusions were avoided. We want to note that the task to detect a CU means that the system has to assign at least one CU for each text, which is necessary to start building a RS-tree. Taking this into account we have compared the systems into 2 groups:

- RB and system's with Post1 that may not return any CU for each text: The best model is LF+SMO+Post1 which provides 0.76 in cross-validation and 0.657 in test. Table 3 shows that LF+SMO+Post1 system is better than rule-based system in 0.048 points in cross-validation and 0.104 points in test dataset. In the second position is AF+MNB+Post1 model which is better than rule-based system and LF+BNB+Post1 system.
- Baseline and system's with Post2 that return at least one CU for each text: The best model is LF+SMO+Post2 which provides 0.728 in cross-validation and 0.644 in test. We can observe that LF+SMO+Post2 system is better than baseline system in 0.031 points in cross-validation and 0.111 points in test dataset. In the second position is AF+MNB+Post2 model which is better than baseline and LF+BNB+Post2 system.

To conclude, if we compare the system's with Post1 and Post2, the task of return at least one CU for each text reduces the accuracy in almost all the systems.

5.1 A comparison using box plot

To show how robust all the algorithms are on the dataset we run 10-fold cross-validation 10 times. The training dataset was randomly broken into 10 partitions using 10 random seeds. We have calculated 10 means of the F-score value for each 10-fold cross-validation (see Figure 2).

To visualize the performance of the 14 systems (Baseline (B), Rule-based (RB),[7] LF+BNB (LBNB), LF+BNB+Post1 (LBNBP1), LF+BNB+Post2 (LBNBP2), LF+SMO (LSMO), LF+SMO+Post1 (LSMOP1), LF+SMO+Post2 (LSMOP2), AF+MNB (AMNB), AF+MNB+Post1 (AMNBP1), AF+MNB+Post2 (AMNBP2), AF+SMO (ASMO), AF+SMO+Post1 (ASMOP1), AF+SMO+Post2 (ASMOP2)), we

[6] Where the nearer it is from 1 the easier it is to determine the CU.

[7] Results obtained in previous work using a rule-based system was 0.712 for train and 0.553 for test.

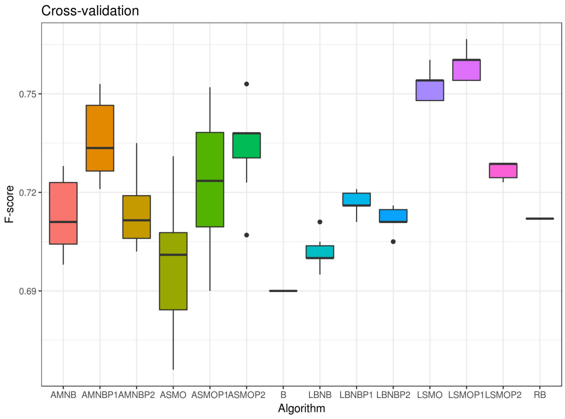

Figure 2: Exploring F-score distribution on the 10-fold cross-validation using 10 random seeds with Box Plot

have summarized the distribution of F-score values using box plots. A box plot consists of a box summarizing 50% of the data. The upper and lower ends of the box are the upper and lower quartiles, while a thick line within the box encodes the median. Dashed appendages summarize the spread and shape of the distribution, and dots represent outside values or outliers.

Taking this into account we have compared the systems into 2 groups:

- RB and system's with Post1 that may not return any CU for each text: The best model is LF+SMO+Post1 which has a F-score median value of 0.76. All the system's with Post1 show a greater F-score median value than RB. In the second position is a system that use automatic features like AF+MNB+Post1 which has a F-score median value of 0.723.

- Baseline and system's with Post2 that return at least one CU for each text: the best model is AF+SMO+Post2 (which has a F-score median value of 0.738) and it is close to LF+SMO+Post2 (which provides a F-score median value of 0.728).

If the system has to assign at least one CU for each text, which is necessary to start building a RS-tree, we finally have selected LF+SMO+Post2 system to avoid outliers.

To understand these results, we present an error analysis in the following subsection.

Correct CUs		**Wrong CUs due to**	
Total agreem.	Partial agreem.	Wrong Structure	Segment. errors
25	4	10	1

Table 4: SMO's post-processed method error analysis of the test dataset

5.2 Error analysis

First of all, we checked manually if each text follows the patterns specified by Antonio (2015) and we found that the 10-fold cross-validation dataset follows in a better way than the test dataset. 25% of the texts do not have the prototypical characteristics of the CUs in the 10-fold cross-validation dataset, whereas in the test dataset 55% of the texts do not have the prototypical characteristics of the CUs, because the analyzed texts were written by students.

Secondly, in the following error analysis in Table 4, we analyze why the SMO+Post2 system does not detect all the CUs of the 40 texts from the test dataset and, mainly, errors happen with texts which do not present the expected structure for the argumentative answer genre.

6 Conclusions and future work

In this paper we have introduced the first Central Unit (CU) detector based on machine learning techniques for Brazilian argumentative answer texts. The CU detector can be tested at `http://ixa2.si.ehu.eus/rst/tresnak/cu-detector/bp/`.

The results from our work indicate that identifying the CU discourse segment in argumentative answer text is well defined and the optimal set of features to classify the CUs are: title words, segment position, epistemic adverbs, copula verbs and evidential verbs in first person singular.

Detecting the CUs in real exams that have been written by students is difficult because sometimes they do not follow the discourse structure of an argumentative answer text, but of a dissertation. Different features have to be taken into account when we are detecting the CU in different genres.

We conclude that this system ensures, with a post-process stage, that each text has at least one central unit and we obtain better results using machine learning techniques (results with BNB+Post2 were 0.58 and with SMO+Post2 0.644) than using a rule-based approach (0.553) (Iruskieta, Antonio, and Labaka, 2016). We think that there is some room to improve using punctuation information as other features in the post-processing, as we have to work with students argumentative answer texts.

We have shown that a CU detector based on machine learning techniques can be built by easily extracting manually some indicator as in Antonio (2015). We think that some parsers can benefit from this methodological step to find the CU after automatic segmentation and before linking the rhetorical relations of an RS-tree.

The work carried out will be useful if we can provide a fair evaluation of the argumentative answer texts, assigning better grades to those texts which follow the patterns of the CU and giving some indicators or clues to students to write the CU in a better way. We intend to develop different studies of how we can detect the CU in other languages, genres and domains taking into account annotated data and features developed here.

Acknowledgments

This study was carried out within the framework of the following projects: IXA Group: natural language processing (GIU16/16) [UPV/EHU], QUALES KK-2017/00094 (Gobierno Vasco) and TUNER TIN2015-65308-C5-1-R (MINECO/FEDER, UE).

References

Aleixo, P. and T. Pardo. 2008. CSTNews: um córpus de textos jornalísticos anotados segundo a teoria discursiva multidocumento CST (Cross-Document Structure Theory). Technical Report ICMC-USP.

Antonio, J. 2015. Detecting central units in argumentative answer genre: signals that influence annotators' agreement. In *5th Workshop "RST and Discourse Studies" in Actas del XXXI Congreso de la Sociedad Española del Procesamiento del Lenguaje Natural*. SEPLN.

Antonio, J. D. and J. A. Santos. 2014. A estrutura retórica do gênero resposta argumentativa. *Signum: Estudos da Linguagem*, 17(2):193–223.

Bengoetxea, K., A. Atutxa, and M. Iruskieta. 2017. Un detector de la unidad central de un texto basado en técnicas de aprendizaje automático en textos científicos para el euskera. *Procesamiento del Lenguaje Natural*, 58:37–44.

Bengoetxea, K. and M. Iruskieta. 2018. A supervised central unit detector for spanish. *Procesamiento del Lenguaje Natural 60: 29–36. ISSN 1135-5948. DOI 10.26342/2018-60-3.*

Burstein, J. and D. Marcu. 2003. A machine learning approach for identification of thesis and conclusion statements in student essays. *Computers and the Humanities*, 37(4):455–467.

Dall'Aglio-Hattnher, M. 2007. Pesquisas em sintaxe: a abordagem funcionalista da evidencialidade. *Trilhas de Mattoso Câmara e outras trilhas: fonologia, morfologia e sintaxe. Araraquara: Cultura Acadêmica Editora*, 12:103–145.

Iruskieta, M., J. Antonio, and G. Labaka. 2016. Detecting the central units in two different genres and languages: a preliminary study of Brazilian Portuguese and Basque texts. *Procesamiento de Lenguaje Natural*, 56:65–72.

Mann, W. and S. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.

Menegassi, R. J. 2011. A escrita na formação docente inicial: influências da iniciação à pesquisa. *Signum: Estudos da Linguagem*, 14(1):387–419.

Pardo, T. and M. Nunes. 2004. Dizer - um analisador discursivo automático para o português do brasil. In *In Anais do IX Workshop de Teses e Dissertações do Instituto de Ciências Matemáticas e de Computação*, pages 1–3, São Carlos-SP, Brasil. 19 a 20 de Novembro.

Pardo, T., L. Rino, and M. Nunes. 2003. GistSumm: A summarization tool based on a new extractive method. *Computational Processing of the Portuguese Language*, pages 196–196.

Platt, J. 1998. Sequential minimal optimization: A fast algorithm for training support vector machines. Technical Report MSR-TR-98-14.