Saying no but meaning yes: negation and sentiment analysis in Basque

Jon Alkorta
Computer Languages
and Systems
IXA group (UPV/EHU)

Koldo Gojenola
Computer Languages
and Systems
IXA group (UPV/EHU)

Mikel Iruskieta
Didactics of Language
and Literature Department
IXA group (UPV/EHU)

`{jon.alkorta,koldo.gojenola,mikel.iruskieta}@ehu.eus`

Abstract

In this work, we have analyzed the effects of negation on the semantic orientation in Basque. The analysis shows that negation markers can strengthen, weaken or have no effect on sentiment orientation of a word or a group of words. Using the Constraint Grammar formalism, we have designed and evaluated a set of linguistic rules to formalize these three phenomena. The results show that two phenomena, strengthening and no change, have been identified accurately and the third one, weakening, with acceptable results.

1 Introduction

Negation is a morphosyntactic operation in which a lexical item denies or inverts the meaning of another lexical item or language construction (Loos et al., 2004). The effect of the negation can be the change of semantic orientation (SO) and, according to Liu (2012), negation is called sentiment shifters because they change the semantic orientation of a word or a sentence.

With the aim of calculating the semantic orientation, the first step is to build a lexicon, but this is not enough, to grasp the correct SO-value of Example 1.

(1) $[Irabazi_{+2}$ _ezinik jarraitzen du Eibarrek$]^{+2}$._ (KIR17)
[(The soccer team) Eibar continues <u>without winning$_{+2}]^{+2}$.</u>

Following the semantic lexicon Sentitegi (Alkorta et al., 2018)[1], the semantic orientation of the word _irabazi_ ("to win") is $+2$, and consequently, of the sentence also is $+2$. But we can notice that the semantic orientation of the sentence is clearly negative. The negator _ezin_ ("can not") turns the positive oriented word $irabazi_{+2}$ ("to win") into a negative oriented one. Therefore, we think that addressing this phenomenon is crucial to obtain better results in the calculation of the SO of texts.

The main aim of this work is to study how negation expressions and syntactic structures can change the semantic orientation of words, and to design a set of linguistic rules by means of Constraint Grammar (Karlsson et al., 2011) in order to identify these phenomena. According to our corpus study, different negation language forms can strengthen, weaken or have no effect on semantic orientation. These results go in the same direction as (Jiménez-Zafra et al., 2018b) where effects of negation within its scope are studied. We have centered our study on negation markers that unlike negation in verbs and nouns and negative polarity items, they only share information about negativity while others can share more information like aspect of action (e.g. _they denied going to the city_).

This paper has been organized as follows: after presenting related work in Section 2, Section 3 describes methodological steps. Then, Section 4 presents theoretical framework, while Section 5 gives a linguistic analysis. Section 6 shows results and error analysis, concluding with Section 7 and proposing directions for future work.

2 Related Work

There is a variety of works about negation and sentiment analysis in different languages and from different approaches.

For English, Liu and Seneff (2009) have presented a work where a parse-and-paraphrase paradigm is used to assign sentiment polarity for product reviews. If negation is detected, its polarity will be reversed (switch negation). If it has a value of $+5$, it will be reversed to -5, and vice versa. Following this, they have improved results (recall was improved in 45 %). The treatment of negation has been different in Taboada et al. (2011). In their work, when a negator is identified, the polarity value is not reversed; instead it is shifted toward the opposite polarity by a fixed amount. This approach is called shift negation. In

---

[1] The semantic lexicon is available on the web at: `http://ixa.si.ehu.es/node/11438`

Text	Text span	Dictionary words	SO value
MUS20	*Pogostkinak [ezin hobeki]⁺ atera zituen*	*hobeki*	+2
	Pogostkina took them out [in an unbeatable way]⁺	best, better	
KIR17	*[Irabazi ezinik]⁻ jarraitzen du Eibarrek,*	*Irabazi*	+2
	Eibar continues [without winning]⁻	to win	

Table 1: Polarity extraction of words (step 2) and linguistic analysis (step 3).

the creation of the semantic orientation calculator (SO-CAL tool), Taboada et al. (2011) have also treated negation in combination with other linguistic phenomena (like irrealis or intensifiers).

In Spanish, there are several works related to negation and sentiment analysis. In the case of Jiménez Zafra et al. (2015), firstly, they have analyzed what the effects of different negators in different sentences are. After that, they have created linguistic rules defined by the previous analysis. Finally, they developed a module that has been included in their polarity classifier system, improving results between 2.25 % and 3.02 % depending on the resource. Vilares et al. (2015) have used a syntactic approach for opinion mining on Spanish reviews. This system treats negation taking into account the scope and polarity flip caused by negation. According to their results, there is an improvement, due to the implementation of negation, among other reasons.

Our work is related to (Taboada et al., 2011) and (Jiménez Zafra et al., 2015) since it is based on a linguistic analysis and also because a set of rules that detect the negation language forms are created. As far as we know, there is not any work which analyzes negation in connection with sentiment analysis in Basque.[2]

3 Methodological steps

1- Negation corpus. We have extracted 359 negation instances of seven[3] negation markers. They were extracted from a total of 96 reviews of six different topics: movies, music, literature, politics, sports and forecast. We have selected those negation markers because they are the most frequent in the corpus.

2- Polarity extraction of every instance. We have created a polarity tagger, based on a POS tagger (Ezeiza et al., 1998) to enrich the corpus with POS information on a se-

mantic oriented lexicon for Basque (Alkorta et al., 2018), to assign the semantic orientation value (SO value, between -5 and $+5$) to words, as shown in Table 1. There, the adverb *hobeki* ("best", "better") and the verb *irabazi* ("to win"), have a SO value of $+2$ in the lexicon.

3- Linguistic analysis. We have analyzed whether the negation markers can change the semantic orientation and the SO value of sentences. We have also tried to identify whether there are other phenomena related to negation with or without effects on semantic orientation. In Table 1, in MUS20, the negation marker appears near *hobeki* ("best"), an adverb. The result of this combination is strengthening. In contrast, in KIR17, the verb *irabazi* ("to win") is before the negator and the result is weakening. These two examples show the different performances of *ezin(ik)* ("can not"). Consequently, in Table 1, for example, this negation marker appears in two different groups. The same methodology has been used with other negation markers.

4- Constraint Grammar (CG3) rules for negation. Several rules have been proposed to detect each group, in order to identify the effects of negation based on the linguistic analysis presented in Section 5.

5- Evaluation. We use F_1 to evaluate the results using a different set of 46 reviews from the same corpus (Alkorta et al., 2016)[4].

4 Theoretical framework

In this section, we explain the three most important concepts, regarding our analysis: i) scope (negation analysis) and ii) switch and iii) shift negation (sentiment analysis approach to negation).

(2) *Berez pianorako konposatutako poliptiko txiki honek ez du bere naf kutsua galtzen$_{-2}$ bertsio orkestratuan.* (MUS01)
This small polyptych composed for the piano does **not** lose$_{-2}$ its naive sense in the orchestral version.

---

[2] Altuna et al. (2017) also analyze negation but their point of view is different, since they analyze events in Basque texts.

[3] The extracted negation markers from the Basque Opinion Corpus (Alkorta et al., 2016) are the following: *ez* ("not"), *gabe* ("without"), *ezin* ("can not"), *salbu* ("except (for)"), *izan ezik* ("except (for)"), *ezta* ("not even, not either") and *ezean* ("in the absence of" or "unless").

[4] A part of the corpus is available on the web at: http://ixa2.si.ehu.es/diskurtsoa/ fitxategiak.php

(3) −*maitasun istorio konbentzional bat, grazia*$_{+3}$ *handirik*$_{+1}$ **gabe***a*−. (LIB07)
−a conventional love story, **without** <u>great$_{+1}$ grace$_{+3}$</u>.[5]

According to Huddleston and Pullum (2002), the scope of negation is the part of the meaning that is affected by the negation marker, changing or not their SO value. In the examples above, the scope is underlined. As our study shows, there can be two kinds of semantic orientation in scope and these can be changed by negation markers. In Example 2, the SO value of the verb *galdu* ("to lose") and of its scope is −2. The negation weakens the SO value of the verb, reversing its SO. But, in Example 3, the SO values of the noun *grazia* ("grace") +3 and the adjective *handi* ("great") +1 assign a SO value of +4 to the scope which is positive. The negator *gabe* ("without") weakens the SO value.

According to Taboada et al. (2011), there are two approaches in sentiment analysis to weaken the negative SO value: $i)$ switch negation and $ii)$ shift negation.

(4) <u>This pub is [**not** <u>good</u>$_{+3}$]$^{-3(switch)}_{-1(shift)}$</u> but the music from there is good$_{+3}$.

In the switch negation approach, the SO value of Example 4 is reversed. The SO value of the adjective *good* is +3 while the reversed SO value is −3. However, this criteria has a problem: if *excellent* is +5; *not excellent* would be more positive (+1) than *not good* (−2), but the SO value points to the contrary (*not excellent* is more negative than *not good*).

Otherwise, in the shift negation, the different negators have their own SO value and the results depend on the interaction of both SO values (the value of negation marker and negated word). Taking into account Example 4, the SO value of the negation *no* is −4 in the dictionary; so, when it modifies the word *good*, which has a SO value of +3, the sum value of scope is −1. This is the way how the shift approach solves the problem we describe in Example 4. We have decided to use the shift negation approach assigning a ±4 SO value to the negators.

5 Linguistic analysis

In the theoretical framework of the shift negation, it has been considered that negation

---

[5]**Bold** is used to mark the negator, <u>underline</u> means the scope of negation.

markers only weakens the SO value. Nevertheless, we have identified two other functions of these negation markers with low frequency, but relevant anyway from our point of view as the works of (Jiménez-Zafra et al., 2018a) and (Jiménez-Zafra et al., 2018b) show. As we observed in this study, the negation markers can strengthen, weaken or have no effect in the SO value of its scope as Figure 1 shows.
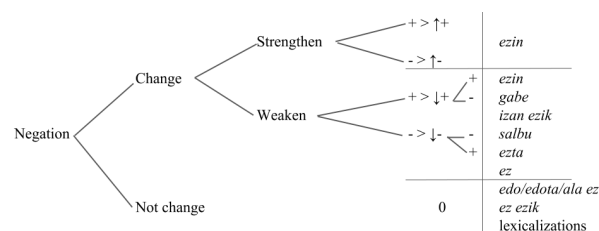

Figure 1: The effects of negation on semantic orientation according to negation markers.

The majority of negation markers usually weaken the semantic orientation of scope. But as we can see in Figure 1, the negation marker *ezin* ("can not"), for example, can strengthen or weaken the semantic orientation of scope. The weakening can be understood in two ways: $i)$ if the word or scope of the semantic orientation is +5, +4, −5 or −4, their semantic orientation will not become negative because according to our methodology (shift negation), due to our SO value of the negators is ±4. In contrast, $ii)$ if the semantic orientation of scope or sentence is between −3 and +3, their semantic orientation will be reversed. $iii)$ Finally, negation with conjunction, contrastive negation and lexicalized structures do not change the SO value of the scope.

5.1 Negation strengthening the SO

Among all the negation instances, we have observed some cases where the semantic orientation has been strengthened (1.96 %: 7 of 359). This happens when the negation marker *ezin* ("can not") modifies adjectives or adverbs.

(5) *Dena nahasten da maisulan* **ezin** *ederragoa*$_{(+4)}$ *osatzeko*. (MUS21)
<u>Everything is mixed to create a masterpiece that can **not** be more beautiful$_{(+4)}$</u>.

In Example 5, the negator modifies the adjective and, in this case, the negation with an adjective in a comparative structure is used to reinforce the positive SO value. The result

Example	Negation marker	Categorization	Instances
6		[(NP +)] *ez* [+ aux. (+ NP) + verb (+ NP)]	214
	ez	[(NP +) verb +] *ez*	18
		[NP +] *ez*	13
		ez [+ NP +] *ez* [+ NP] (...) (repetitive)	2
	gabe	[NP/VP/clause +] *gabe*	41
7	*ezin*	[(NP) + verb +] *ezin*	19
		ezin [(+ NP +)] verb [(+ NP)]	5
	salbu	[NP] + *salbu*	2
	izan ezik	[NP/clause] + *izan ezik*	1
	ezta	*ezta* + [NP/clause]	1
	ez, ezin	with any clear pattern	7
	Total		**323**

Table 2: Negation weakening the semantic orientation.

of negating a positive chunk *can not be more beautiful* is to be even more positive. In this case, the *masterpiece* is *very beautiful*.

5.2 Negation weakening the SO

In the majority of cases, the SO value is weakened due to negation. Several negation markers can weaken the semantic orientation. In our corpus, 89.98 % of cases (323 of 359) show a weakening of scope.

(6) *Horrek* **ez** *die eragotzi*$_{(-2)}$ *ordea, 57 milioi euro ematea San Mames klub pribatuari!.* (POL30)
It does **not** prevent$_{(-2)}$ them, however, to give 57 milion euros to San Mames private club!

(7) *Irabazi*$_{(+2)}$ ***ezin****ik jarraitzen du Eibarrek, baina oso puntu ona eskuratu du Getaferen zelaian.* (KIR17)
Eibar continues **without** winning$_{(+2)}$, but it has achieved a very good point in Getafe's (football) field.

In Example 6, the default word order of Basque (main verb + auxiliary verb) was reversed in a typical negation structure (*ez* "not" + auxiliary verb + main verb). In this example, the negation marker *ez* ("not") has an effect on all the words of the sentence, including the verb *eragotzi* ("prevent") which has a negative SO value (−2), weakening its SO value. In Example 7, the negation marker *ezin* "can not" negates the verb *irabazi* ("win"). Therefore, the negation marker *ezin* ("can not") works like an intensifier does with adjectives and adverbs (Example 5) while it has the opposite function with verbs and nouns (Example 7). Therefore, weakening negators can have a positive or negative (±4) SO value, if the modified chunk (scope) has a positive or negative SO value. The same happens if the SO value is positive +5, because the result of the weakening (−4) will not change the polarity and

the SO value will still be positive +1. In contrast, if the SO value of the modified chunk +3 or −3 or lower, the SO value will be reversed to a ±1. This happens in Example 6 and Example 7. In the first example, the SO value of the scope is +2 (*eragotzi* ("prevent") −2 + *ez* ("not") +4 = +2). In the second one, the SO value of the scope is −2 (*irabazi* ("win") +2 + *ez* ("not") −4 = −2).

5.3 Negation with no effect

Negation with no effect on semantic orientation has happened in 8.08 % of our sample (27 of 359). In these cases, the negation does not modify any word with a SO value assigned. This can happen due to three reasons: *i*) the negator appears with a conjunction, *ii*) the negator is a part of contrastive negation and *iii*) the negator is part of a lexicalized structure (structures with their own meaning and sometimes also corresponding to dictionary entries). The scope concept is applicable only in the case of contrastive negation and the particle *ez* ("no") with a conjunction.

(8) *Ikuspuntu politikotik*$_{(-1)}$ *ez ezik, ekonomiko-tik*$_{(+3)}$ *ere Greziak esperantza ekarri du Europako hegoaldeko beste herrietara, tartean Euskal Herrira.* (POL08)
Not only from the political point of view, but also from the economic point of view, Greece has also hoped for other parts of southern Europe, including the Basque Country.

(9) *Sei puntu baino ez dituela, hamaseigarren postuan da Reala sailkapenean.* (KIR27)
With only six points, Real is in the sixteenth position in the classification.

Example 8 shows a contrastive negation with additive function (Silvennoinen, 2017). In other words, the negation mark does not negate the noun phrase, as in *ikuspuntu politikotik*$_{(-1)}$ ("from the political$_{(-1)}$ point of view"), actually it functions as conjunction

Example	Negation marker / lexicalized structure	Instances
	[verb/*bai* "yes"] + *edo/edota/ala ez* (*ez* with conjuction)	3
8	[NP] + *ez ezik* (contrastive negation)	2
9	*baino/besterik ez*	11
	Others lexicalized structures	13
	Total	**29**

Table 3: Negation without effects on semantic orientation.

and adds new information: *ekonomikotik ere* ("also from the economic point of view"). Structures of Table 3 have their own SO value, they can be considered as dictionary entries and they can appear in different positions in the sentence. In Example 9, the structure *baino/besterik ez* ("only") is an adverb.

6 Evaluation
6.1 Evaluation methodology
To tag the negation changes of the SO value, we have created negation rules based on previous studies. Rules have been implemented using Constraint Grammar (CG3) (Karlsson et al., 2011) to assign the correct value to the negated structures. The corpus of 96 texts has been tagged using the Basque morphosyntactic disambiguator based on the CG formalism (Aduriz et al., 1997). Then, a different set of 48 texts of the Basque Opinion Corpus has been used as test dataset to evaluate the rules. After that, the results have been analyzed manually, observing if the words have been annotated or not and, when annotated, whether they have the correct annotation.

Negation effects	Prec.	Rec.	F_1
Strengthen	1.00	1.00	1.00
Weaken	0.93	0.80	0.86
No effect	0.97	1.00	0.98
Total	**0.93**	**0.80**	**0.86**
Negated elements	**Prec.**	**Rec.**	**F_1**
Negation markers	1.00	0.96	0.98
Lexicalized structures	0.96	1.00	0.98
Scope	0.91	0.75	0.82
Total	**0.93**	**0.80**	**0.86**

Table 4: General results of negation effects and negated elements.

Most of the corpus was evaluated by one linguist, but with the aim to know the reliability of this evaluation a piece of the corpus (10 %) has been annotated by two linguists. Both annotators have followed a guideline to evaluate the output of CG3 rules. According to the results, the Cohen's kappa score is 0.93 for the annotation of the words that belong to negation and the kappa score is 0.69 for the annotation of words that have been annotated correctly, badly or is missed (which can be considered as *substantial* in (Landis and Koch, 1977)).

6.2 Results and error analysis
According to general results, the F_1 of the negation rules identifying elements related to negation is 0.86 (Precision is 0.93 while recall is 0.80).

In accordance with weakening and scope error analysis, these elements show lower F_1 score because they behave more irregularly. The components as well as the length in scope are more unpredictable. Moreover, some negators apply to lists of words with comma and, as some constraints in CG3 rules correspond to punctuation marks, they have not been detected. This suggests that the rules need more precision. So, the punctuation mark constraint is not enough. Therefore, some syntactic information is needed to detect these kind of structures.

7 Conclusions and Future Work
This work presents a negation analysis for Basque sentiment analysis based on Constraint Grammar rules. According to this study, the negation can affect the semantic orientation (SO value) in different ways: i) strengthening, ii) weakening or iii) having no effect. According to our evaluation to measure the identified words, the overall precision is 0.93, the recall 0.80 and the F_1 score 0.86. In line with error analysis, the punctuation mark constraint is not enough and more precise rules are needed in the negation weakening. In the near future, i) we want to implement these negation rules in a tool for automatic Basque sentiment analysis and ii) we want to continue with the analysis of negation: analyzing the scope in a bigger corpus and especially based on the Rhetorical Structure Theory (RST) (Mann and Thompson, 1987), studying if the position of negator in rhetorical structure has any effect on sentiment analysis.

References

Itziar Aduriz, José María Arriola, Xabier Artola, Arantza Díaz de Ilarraza, Koldo Gojenola, and Montse Maritxalar. 1997. Morphosyntactic disambiguation for basque based on the constraint grammar formalism. *Proceedings of Recent Advances in NLP (RANLP97)*, pages 282–288, Tzigov Chark (Bulgaria).

Jon Alkorta, Koldo Gojenola, and Mikel Iruskieta. 2016. Creating and evaluating a polarity - balanced corpus for Basque sentiment analysis. In *IWoDA16 Fourth International Workshop on Discourse Analysis*, pages 58–62. Santiago de Compostela (Spain).

Jon Alkorta, Koldo Gojenola, and Mikel Iruskieta. 2018. SentiTegi: building a semantic oriented Basque lexicon. In *Proceedings of the CICLing 2018*. Hanoi (Vietnam).

María Jesus Aranzabe Begoña Altuna and Arantza Díaz de Ilarraza. 2017. Euskarazko ezeztapenaren tratamendu automatikorako azterketa. In *Iñaki Alegria, Ainhoa Latatu, Miren Josu Ormaetxebarria and Patxi Salaberri (pub.), II. IkerGazte, Nazioarteko Ikerketa Euskaraz: Giza Zientziak eta Artea*, pages 127–134. Udako Euskal Unibertsitatea (UEU), Bilbo (Spain).

Nerea Ezeiza, Iñaki Alegria, José María Arriola, Rubén Urizar, and Itziar Aduriz. 1998. Combining stochastic and rule-based methods for disambiguation in agglutinative languages. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 380–384. Association for Computational Linguistics.

Rodney Huddleston and Geoffrey Keith Pullum. 2002. The Cambridge grammar of English. *Language. Cambridge: Cambridge University Press*.

Salud María Jiménez-Zafra, M. Teresa Martín-Valdivia, M. Dolores Molina-González, and L. Alfonso Ureña-López. 2018a. Relevance of the SFU Review SP-NEG corpus annotated with the scope of negation for supervised polarity classification in Spanish. *Information Processing & Management*, 54(2):240–251.

Salud María Jiménez Zafra, Eugenio Martínez Cámara, María Teresa Martín Valdivia, and María Dolores Molina González. 2015. Tratamiento de la Negación en el Análisis de Opiniones en Espanol. *Procesamiento del Lenguaje Natural*, (54).

Salud María Jiménez-Zafra, Mariona Taulé, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, and M. Antónia Martí. 2018b. SFU Review SP-NEG: a Spanish corpus annotated with negation for sentiment analysis. A typology of negation patterns. *Language Resources and Evaluation*, 52(2):533–569.

Fred Karlsson, Atro Voutilainen, Juha Heikkilae, and Arto Anttila. 2011. *Constraint Grammar: a language-independent system for parsing unrestricted text*, volume 4. Walter de Gruyter.

J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Jingjing Liu and Stephanie Seneff. 2009. Review sentiment scoring via a parse-and-paraphrase paradigm. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 161–169. Association for Computational Linguistics.

Eugene Emil Loos, Susan Anderson, Dwight H. Day, Paul C. Jordan, and J. Douglas Wingate. 2004. *Glossary of linguistic terms*, volume 29. SIL International Camp Wisdom Road Dallas.

William C. Mann and Sandra A. Thompson. 1987. *Rhetorical Structure Theory: A theory of text organization*. University of Southern California, Information Sciences Institute.

Olli O. Silvennoinen. 2017. Not only apples but also oranges: Contrastive negation and register. *In Turo Hiltunen, Joe McVeigh and Tanja Sily (edit.), Big and Rich Data in English Corpus Linguistics: Methods and Explorations, VARIENG, Helsinki (Finland)*.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

David Vilares, Miguel A. Alonso, and Carlos Gómez-Rodríguez. 2015. A syntactic approach for opinion mining on Spanish reviews. *Natural Language Engineering*, 21(1):139–163.