Semantic Representation of Negation Using Focus Detection

Eduardo Blanco and **Dan Moldovan**
Human Language Technology Research Institute
The University of Texas at Dallas
Richardson, TX 75080 USA
{eduardo,moldovan}@hlt.utdallas.edu

Abstract

Negation is present in all human languages and it is used to reverse the polarity of part of statements that are otherwise affirmative by default. A negated statement often carries positive implicit meaning, but to pinpoint the positive part from the negative part is rather difficult. This paper aims at thoroughly representing the semantics of negation by revealing implicit positive meaning. The proposed representation relies on focus of negation detection. For this, new annotation over PropBank and a learning algorithm are proposed.

1 Introduction

Understanding the meaning of text is a long term goal in the natural language processing community. Whereas philosophers and linguists have proposed several theories, along with models to represent the meaning of text, the field of computational linguistics is still far from doing this automatically. The ambiguity of language, the need to detect implicit knowledge, and the demand for commonsense knowledge and reasoning are a few of the difficulties to overcome. Substantial progress has been made, though, especially on detection of semantic relations, ontologies and reasoning methods.

Negation is present in all languages and it is always the case that statements are affirmative by default. Negation is marked and it typically signals something unusual or an exception. It may be present in all units of language, e.g., words (_incredible_), clauses (_He doesn't have friends_). Negation and its correlates (truth values, lying,

irony, false or contradictory statements) are exclusive characteristics of humans (Horn, 1989; Horn and Kato, 2000).

Negation is fairly well-understood in grammars; the valid ways to express a negation are documented. However, there has not been extensive research on detecting it, and more importantly, on representing the semantics of negation. Negation has been largely ignored within the area of semantic relations.

At first glance, one would think that interpreting negation could be reduced to finding negative keywords, detect their scope using syntactic analysis and reverse its polarity. Actually, it is more complex. Negation plays a remarkable role in text understanding and it poses considerable challenges.

Detecting the scope of negation in itself is challenging: _All vegetarians do not eat meat_ means that vegetarians do not eat meat and yet _All that glitters is not gold_ means that it is not the case that all that glitters is gold (so out of all things that glitter, some are gold and some are not). In the former example, the universal quantifier _all_ has scope over the negation; in the latter, the negation has scope over _all_.

In logic, two negatives always cancel each other out. On the other hand, in language this is only theoretically the case: _she is not unhappy_ does not mean that _she is happy_; it means that _she is not fully unhappy, but she is not happy either_.

Some negated statements carry a positive implicit meaning. For example, _cows do not eat meat_ implies that _cows eat something other than meat_. Otherwise, the speaker would have stated _cows do not eat_. A clearer example is the correct and yet puzzling statement _tables do not eat meat_. This sentence sounds

581

unnatural because of the underlying positive statement (i.e., *tables eat something other than meat*).

Negation can express *less than* or *in between* when used in a scalar context. For example, *John does not have three children* probably means that he has either one or two children. Contrasts may use negation to disagree about a statement and not to negate it, e.g., *That place is not big, it is massive* defines the place as *massive*, and therefore, *big*.

2   Related Work

Negation has been widely studied outside of computational linguistics. In logic, negation is usually the simplest unary operator and it reverses the truth value. The seminal work by Horn (1989) presents the main thoughts in philosophy and psychology. Linguists have found negation a complex phenomenon; Huddleston and Pullum (2002) dedicate over 60 pages to it. Negation interacts with quantifiers and anaphora (Hintikka, 2002), and influences reasoning (Dowty, 1994; Sánchez Valencia, 1991). Zeijlstra (2007) analyzes the position and form of negative elements and negative concords.

Rooth (1985) presented a theory of focus in his dissertation and posterior publications (e.g., Rooth (1992)). In this paper, we follow the insights on scope and focus of negation by Huddleston and Pullum (2002) rather than Rooth's (1985).

Within natural language processing, negation has drawn attention mainly in sentiment analysis (Wilson et al., 2009; Wiegand et al., 2010) and the biomedical domain. Recently, the Negation and Speculation in NLP Workshop (Morante and Sporleder, 2010) and the CoNLL-2010 Shared Task (Farkas et al., 2010) targeted negation mostly on those subfields. Morante and Daelemans (2009) and Özgür and Radev (2009) propose scope detectors using the BioScope corpus. Councill et al. (2010) present a supervised scope detector using their own annotation. Some NLP applications deal indirectly with negation, e.g., machine translation (van Munster, 1988), text classification (Rose et al., 2003) and recognizing entailments (Bos and Markert, 2005).

Regarding corpora, the BioScope corpus annotates negation marks and linguistic scopes exclusively on biomedical texts. It does not annotate focus and it purposely ignores negations such as *(talk-*

ing about the reaction of certain elements) in NK3.3 cells is not always identical (Vincze et al., 2008), which carry the kind of positive meaning this work aims at extracting (in NK3.3 cells *is often* identical). PropBank (Palmer et al., 2005) only indicates the verb to which a negation mark attaches; it does not provide any information about the scope or focus. FrameNet (Baker et al., 1998) does not consider negation and FactBank (Saurí and Pustejovsky, 2009) only annotates degrees of factuality for events.

None of the above references aim at detecting or annotating the focus of negation in natural language. Neither do they aim at carefully representing the meaning of negated statements nor extracting implicit positive meaning from them.

3   Negation in Natural Language

Simply put, negation is a process that turns a statement into its opposite. Unlike affirmative statements, negation is marked by words (e.g., *not*, *no*, *never*) or affixes (e.g., *-n't*, *un-*). Negation can interact with other words in special ways. For example, negated clauses use different connective adjuncts that positive clauses do: *neither*, *nor* instead of *either*, *or*. The so-called *negatively-oriented polarity-sensitive items* (Huddleston and Pullum, 2002) include, among many others, words starting with *any-* (*anybody*, *anyone*, *anywhere*, etc.), the modal auxiliaries *dare* and *need* and the grammatical units *at all*, *much* and *till*. Negation in verbs usually requires an auxiliary; if none is present, the auxiliary *do* is inserted (*I read the paper* vs. *I didn't read the paper*).

3.1   Meaning of Negated Statements

State-of-the-art semantic role labelers (e.g., the ones trained over PropBank) do not completely represent the meaning of negated statements. Given *John didn't build a house to impress Mary*, they encode AGENT(*John*, *build*), THEME(*a house*, *build*), PURPOSE(*to impress Mary*, *build*), NEGATION(*n't*, *build*). This representation corresponds to the interpretation *it is not the case that John built a house to impress Mary*, ignoring that it is implicitly stated that *John did build a house*.

Several examples are shown Table 1. For all statements *s*, current role labelers would only encode *it is not the case that* s. However, examples (1–7)

	Statement	Interpretation
1	John didn't build a house to impress Mary.	John built a house for other purpose.
2	I don't have a watch with me.	I have a watch, but it is not with me.
3	We don't have an evacuation plan for flooding.	We have an evacuation plan for something else (e.g., fire).
4	They didn't release the UFO files until 2008.	They released the UFO files in 2008.
5	John doesn't know exactly how they met.	John knows how they met, but not exactly.
6	His new job doesn't require driving.	His new job has requirements, but it does not require driving.
7	His new job doesn't require driving yet.	His new job requires driving in the future.
8	His new job doesn't require anything.	His new job has no requirements.
9	A panic on Wall Street doesn't exactly inspire confidence.	A panic on Wall Streen discourages confidence.

Table 1: Examples of negated statements and their interpretations considering underlying positive meaning. A wavy underline indicates the focus of negation (Section 3.3); examples (8, 9) do not carry any positive meaning.

carry positive meaning underneath the direct meaning. Regarding (4), encoding that the UFO files *were released in 2008* is crucial to fully interpret the statement. (6–8) show that different verb arguments modify the interpretation and even signal the existence of positive meaning. Examples (5, 9) further illustrate the difficulty of the task; they are very similar (both have AGENT, THEME and MANNER) and their interpretation is altogether different. Note that (8, 9) do not carry any positive meaning; even though their interpretations do not contain a verbal negation, the meaning remains negative. Some examples could be interpreted differently depending on the context (Section 4.2.1).

This paper aims at thoroughly representing the semantics of negation by revealing implicit positive meaning. The main contributions are: (1) interpretation of negation using focus detection; (2) focus of negation annotation over all PropBank negated sentences[1]; (3) feature set to detect the focus of negation; and (4) model to semantically represent negation and reveal its underlying positive meaning.

3.2 Negation Types

Huddleston and Pullum (2002) distinguish four contrasts for negation:

- Verbal if the marker of negation is grammatically associated with the verb (*I did not see anything at all*); non-verbal if it is associated with a dependent of the verb (*I saw nothing at all*).
- Analytic if the sole function of the negated mark is to mark negation (*Bill did not go*); synthetic if it has some other function as well (*[Nobody]*AGENT *went to the meeting*).

[1] Annotation will be available on the author's website

- Clausal if the negation yields a negative clause (*She didn't have a large income*); subclausal otherwise (*She had a not inconsiderable income*).
- Ordinary if it indicates that something is not the case, e.g., (1) *She didn't have lunch with my old man: he couldn't make it*; metalinguistic if it does not dispute the truth but rather reformulates a statement, e.g., (2) *She didn't have lunch with your 'old man': she had lunch with your father*. Note that in (1) the lunch never took place, whereas in (2) a lunch did take place.

In this paper, we focus on verbal, analytic, clausal, and both metalinguistic and ordinary negation.

3.3 Scope and Focus

Negation has both scope and focus and they are extremely important to capture its semantics. Scope is the part of the meaning that is negated. Focus is that part of the scope that is most prominently or explicitly negated (Huddleston and Pullum, 2002).

Both concepts are tightly connected. Scope corresponds to all elements any of whose individual falsity would make the negated statement true. Focus is the element of the scope that is *intended* to be interpreted as false to make the overall negative true.

Consider (1) *Cows don't eat meat* and its positive counterpart (2) *Cows eat meat*. The truth conditions of (2) are: (a) somebody eats something; (b) cows are the ones who eat; and (c) meat is what is eaten.

In order for (2) to be true, (a–c) have to be true. And the falsity of any of them is sufficient to make (1) true. In other words, (1) would be true if *nobody eats*, *cows don't eat* or *meat is not eaten*. Therefore, all three statements (a–c) are inside the scope of (1).

The focus is more difficult to identify, especially

1	AGENT(*the cow*, *didn't eat*)	THEME(*grass*, *didn't eat*)	INSTRUMENT(*with a fork*, *didn't eat*)
2	NOT[AGENT(*the cow*, *ate*)	THEME(*grass*, *ate*)	INSTRUMENT(*with a fork*, *ate*)]
3	NOT[AGENT(*the cow*, *ate*)]	THEME(*grass*, *ate*)	INSTRUMENT(*with a fork*, *ate*)
4	AGENT(*the cow*, *ate*)	NOT[THEME(*grass*, *ate*)]	INSTRUMENT(*with a fork*, *ate*)
5	AGENT(*the cow*, *ate*)	THEME(*grass*, *ate*)	NOT[INSTRUMENT(*with a fork*, *ate*)]

Table 2: Possible semantic representations for *The cow didn't eat grass with a fork*.

without knowing stress or intonation. Text understanding is needed and context plays an important role. The most probable focus for (1) is *meat*, which corresponds to the interpretation *cows eat something else than meat*. Another possible focus is *cows*, which yields *someone eats meat, but not cows*.

Both scope and focus are primarily semantic, highly ambiguous and context-dependent. More examples can be found in Tables 1 and 3 and (Huddleston and Pullum, 2002, Chap. 9).

4 Approach to Semantic Representation of Negation

Negation does not stand on its own. To be useful, it should be added as part of another existing knowledge representation. In this Section, we outline how to incorporate negation into semantic relations.

4.1 Semantic Relations

Semantic relations capture connections between concepts and label them according to their nature. It is out of the scope of this paper to define them in depth, establish a set to consider or discuss their detection. Instead, we use generic semantic roles.

Given *s: The cow didn't eat grass with a fork*, typical semantic roles encode AGENT(*the cow*, *eat*), THEME(*grass*, *eat*), INSTRUMENT(*with a fork*, *eat*) and NEGATION(*n't*, *eat*). This representation only differs on the last relation from the positive counterpart. Its interpretation is *it is not the case that* s.

Several options arise to thoroughly represent *s*. First, we find it useful to consider the semantic representation of the affirmative counterpart: AGENT(*the cow*, *ate*), THEME(*grass*, *ate*), and INSTRUMENT(*with a fork*, *ate*). Second, we believe detecting the focus of negation is useful. Even though it is open to discussion, the focus corresponds to INSTRUMENT(*with a fork*, *ate*) Thus, the negated statement should be interpreted as *the cow ate grass, but it did not do so using a fork*.

Table 2 depicts five different possible semantic representations. Option (1) does not incorporate any explicit representation of negation. It attaches the negated mark and auxiliary to *eat*; the negation is part of the relation arguments. This option fails to detect any underlying positive meaning and corresponds to the interpretation *the cow did not eat*, *grass was not eaten* and *a fork was not used to eat*.

Options (2–5) embody negation into the representation with the *pseudo-relation* NOT. NOT takes as its argument an instantiated relation or set of relations and indicates that they do not hold.

Option (2) includes all the scope as the argument of NOT and corresponds to the interpretation *it is not the case that the cow ate grass with a fork*. Like typical semantic roles, option (2) does not reveal the implicit positive meaning carried by statement *s*. Options (3–5) encode different interpretations:

- (3) negates the AGENT; it corresponds to *the cow didn't eat, but grass was eaten with a fork*.
- (4) applies NOT to the THEME; it corresponds to *the cow ate something with a fork, but not grass*.
- (5) denies the INSTRUMENT, encoding the meaning *the cow ate grass, but it did not use a fork*.

Option (5) is preferred since it captures the best implicit positive meaning. It corresponds to the semantic representation of the affirmative counterpart after applying the pseudo-relation NOT over the focus of the negation. This fact justifies and motivates the detection of the focus of negation.

4.2 Annotating the Focus of Negation

Due to the lack of corpora containing annotation for focus of negation, new annotation is needed. An obvious option is to add it to any text collection. However, building on top of publicly available resources is a better approach: they are known by the community, they contain useful information for detecting the focus of negation and tools have already been developed to predict their annotation.

584

	Statement	V	A0	A1	A2	A4	TMP	MNR	ADV	LOC	PNC	EXT	DIS	MOD
1	Even if [that deal]$_{A1}$ isn't [revived]$_V$, NBC hopes to find another.													
	– Even if that deal is suppressed, NBC hopes to find another one.	★	-	+	-	-	-	-	-	-	-	-	-	-
2	[He]$_{A0}$ [simply]$_{MDIS}$ [ca]$_{MMOD}$n't [stomach]$_V$ [the taste of Heinz]$_{A1}$, she says.													
	– He simply can stomach any ketchup but Heinz's.	+	+	★	-	-	-	-	-	-	-	-	+	+
3	[A decision]$_{A1}$ isn't [expected]$_V$ [until some time next year]$_{MTMP}$.													
	– A decision is expected at some time next year.	+	-	+	-	-	★	-	-	-	-	-	-	-
4	[. . .] it told the SEC [it]$_{A0}$ [could]$_{MMOD}$n't [provide]$_V$ [financial statements]$_{A1}$ [by the end of its first extension]$_{MTMP}$ "[without unreasonable burden or expense]$_{MMNR}$".													
	– It could provide them by that time with a huge overhead.	+	+	+	-	-	+	★	-	-	-	-	-	+
5	[For example]$_{MDIS}$, [P&G]$_{A0}$ [up until now]$_{MTMP}$ hasn't [sold]$_V$ [coffee]$_{A1}$ [to airlines]$_{A2}$ and does only limited business with hotels and large restaurant chains.													
	– Up until now, P&G has sold coffee, but not to airlines.	+	+	+	★	-	+	-	-	-	-	-	+	-
6	[Decent life . . .]$_{A1}$ [wo]$_{MMOD}$n't be [restored]$_V$ [unless the government reclaims the streets from the gangs]$_{MADV}$.													
	– It will be restored if the government reclaims the streets from the gangs.	+	-	+	-	-	-	-	★	-	-	-	-	+
7	But [quite a few money managers]$_{A0}$ aren't [buying]$_V$ [it]$_{A1}$.													
	– Very little managers are buying it.	+	★	+	-	-	-	-	-	-	-	-	-	-
8	[When]$_{MTMP}$ [she]$_{A0}$ isn't [performing]$_V$ [for an audience]$_{MPNC}$, she prepares for a song by removing the wad of gum from her mouth, and indicates that she's finished by sticking the gum back in.													
	– She prepares in that way when she is performing, but not for an audience.	+	+	-	-	-	+	-	-	-	★	-	-	-
9	[The company's net worth]$_{A1}$ [can]$_{MMOD}$not [fall]$_V$ [below $185 million]$_{A4}$ [after the dividends are issued]$_{MTMP}$.													
	– It can fall after the dividends are issued, but not below $185 million.	+	-	+	-	★	+	-	-	-	-	-	-	+
10	Mario Gabelli, an expert at spotting takeover candidates, says that [takeovers]$_{A1}$ aren't [totally]$_{MEXT}$ [gone]$_V$.													
	– Mario Gabelli says that takeovers are partially gone.	+	-	+	-	-	-	-	-	-	-	★	-	-

Table 3: Negated statements from PropBank and their interpretation considering underlying positive meaning. Focus is underlined; '+' indicates that the role is present, '-' that it is not and '★' that it corresponds to the focus of negation.

We decided to work over PropBank. Unlike other resources (e.g., FrameNet), gold syntactic trees are available. Compared to the BioScope corpus, Prop-Bank provides semantic annotation and is not limited to the biomedical domain. On top of that, there has been active research on predicting PropBank roles for years. The additional annotation can be readily used by any system trained with PropBank, quickly incorporating interpretation of negation.

4.2.1 Annotation Guidelines

The focus of a negation involving verb v is resolved as:

- If it cannot be inferred that an action v occurred, focus is role MNEG.
- Otherwise, focus is the role that is most prominently negated.

All decisions are made considering as context the previous and next sentence. The mark -NOT is used to indicate the focus. Consider the following statement (file wsj_2282, sentence 16).

[While profitable]$_{MADV_{1,2}}$, [it]$_{A1_1,A0_2}$ "was[n't]$_{MNEG_1}$ [growing]$_{v_1}$ and was[n't]$_{MNEG_2}$ [providing]$_{v_2}$ [a satisfactory return on invested capital]$_{A1_2}$," he says.

The previous sentence is *Applied, then a closely held company, was stagnating under the management of its controlling family*. Regarding the first verb (*growing*), one cannot infer that anything was growing, so focus is MNEG. For the second verb (*providing*), it is implicitly stated that the company was providing *a not satisfactory return on investment*, therefore, focus is A1.

The guidelines assume that the focus corresponds to a single role or the verb. In cases where more than one role could be selected, the most likely focus is chosen; context and text understanding are key. We define the most likely focus as the one that yields the most meaningful implicit information.

For example, in (Table 3, example 2) *[He]*$_{A0}$ could be chosen as focus, yielding *someone can stomach the taste of Heinz, but not him*. However, given the previous sentence (*[. . .] her husband is*
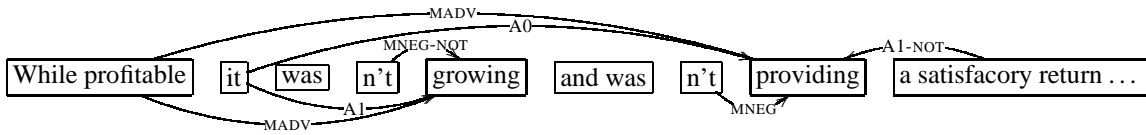
Figure 1: Example of focus annotation (marked with -NOT). Its interpretation is explained in Section 4.2.2.

adamant about eating only Hunt's ketchup), it is clear that the best option is A1. Example (5) has a similar ambiguity between A0 and A2, example (9) between MTMP and A4, etc. The role that yields the most useful positive implicit information given the context is always chosen as focus.

Table 3 provides several examples having as their focus different roles. Example (1) does not carry any positive meaning, the focus is V. In (2–10) the verb must be interpreted as affirmative, as well as all roles except the one marked with '⋆' (i.e., the focus). For each example, we provide PropBank annotation (top), the new annotation (i.e., the focus, bottom right) and its interpretation (bottom left).

4.2.2 Interpretation of -NOT

The mark -NOT is interpreted as follows:

- If MNEG-NOT(*x*, *y*), then verb *y* must be negated; the statement does not carry positive meaning.
- If any other role is marked with -NOT, ROLE-NOT(*x*, *y*) must be interpreted as *it is not the case that x is* ROLE *of y*.

Unmarked roles are interpreted positive; they correspond to implicit positive meaning. Role labels (A0, MTMP, etc.) maintain the same meaning from PropBank (Palmer et al., 2005). MNEG can be ignored since it is overwritten by -NOT.

The new annotation for the example (Figure 1) must be interpreted as: *While profitable, it (the company) was not growing and was providing a not satisfactory return on investment*. Paraphrasing, *While profitable, it was shrinking or idle and was providing an unsatisfactory return on investment*. We discover an entailment and an implicature respectively.

4.3 Annotation Process

We annotated the 3,993 verbal negations signaled with MNEG in PropBank. Before annotation began, all semantic information was removed by mapping all role labels to ARG. This step is necessary to ensure that focus selection is not biased by the seman-

Role	#Inst.	Focus	
		# –	%
A1	2,930	1,194 – 40.75	
MNEG	3,196	1,109 – 34.70	
MTMP	609	246 – 40.39	
MMNR	250	190 – 76.00	
A2	501	179 – 35.73	
MADV	466	94 – 20.17	
A0	2,163	73 – 3.37	
MLOC	114	22 – 19.30	
MEXT	25	22 – 88.00	
A4	26	22 – 84.62	
A3	48	18 – 37.50	
MDIR	35	13 – 37.14	
MPNC	87	9 – 10.34	
MDIS	287	6 – 2.09	

Table 4: Roles, total instantiations and counts corresponding to focus over training and held-out instances.

tic labels provided by PropBank.

As annotation tool, we use Jubilee (Choi et al., 2010). For each instance, annotators decide the focus given the full syntactic tree, as well as the previous and next sentence. A post-processing step incorporates focus annotation to the original PropBank by adding -NOT to the corresponding role.

In a first round, 50% of instances were annotated twice. Inter-annotator agreement was 0.72. After careful examination of the disagreements, they were resolved and annotators were given clearer instructions. The main point of conflict was selecting a focus that yields valid implicit meaning, but not the most valuable (Section 4.2.1). Due to space constraints, we cannot elaborate more on this issue. The remaining instances were annotated once. Table 4 depicts counts for each role.

5 Learning Algorithm

We propose a supervised learning approach. Each sentence from PropBank containing a verbal negation becomes an instance. The decision to be made is to choose the role that corresponds to the focus.

No.	Feature	Values	Explanation
1	role-present	{y, n}	is role present?
2	role-f-pos	{DT, NNP, ...}	First POS tag of role
3	role-f-word	{This, to, overseas, ...}	First word of role
4	role-length	$\mathbb{N}$	number fo words in role
5	role-posit	$\mathbb{N}$	position within the set of roles
6	A1-top	{NP, SBAR, PP, ...}	syntactic node of A1
7	A1-postag	{y, n}	does A1 contain the tag *postag*?
8	A1-keyword	{y, n}	does A1 cotain the word *keyword*?
9	first-role	{A1, MLOC, ...}	label of the first role
10	last-role	{A1, MLOC, ...}	label of the last role
11	verb-word	{appear, describe, ...}	main verb
12	verb-postag	{VBN, VBZ, ...}	POS tag main verb
13	VP-words	{were-n't, be-quickly, ...}	sequence of words of VP until `verb`
14	VP-postags	{VBP-RB-RB-VBG, VBN-VBG, ...}	sequence of POS tags of VP until `verb`
15	VP-has-CC	{y, n}	does the VP contain a CC?
16	VP-has-RB	{y, n}	does the VP contain a RB?
17	predicate	{rule-out, come-up, ...}	predicate
18	them-role-A0	{preparer, assigner, ...}	thematic role for A0
19	them-role-A1	{effort, container, ...}	thematic role for A1
20	them-role-A2	{audience, loaner, ...}	thematic role for A2
21	them-role-A3	{intensifier, collateral, ...}	thematic role for A3
22	them-role-A4	{beneficiary, end point, ...}	thematic role for A4

Table 5: Full set of features. Features (1–5) are extracted for all roles, (7, 8) for all POS tags and keywords detected.

The 3,993 annotated instances are divided into training (70%), held-out (10%) and test (20%). The held-out portion is used to tune the feature set and results are reported for the test split only, i.e., using unseen instances. Because PropBank adds semantic role annotation on top of the Penn TreeBank, we have available syntactic annotation and semantic role labels for all instances.

5.1 Baselines

We implemented four baselines to measure the difficulty of the task:

- A1: select A1, if not present then MNEG.
- FIRST: select first role.
- LAST: select last role.
- BASIC: same than FOC-DET but only using features *last_role* and flags indicating the presence of roles.

5.2 Selecting Features

The BASIC baseline obtains a respectable accuracy of 61.38 (Table 6). Most errors correspond to instances having as focus the two most likely foci: A1

and MNEG (Table 4). We improve BASIC with an extended feature set which targets especially A1 and the verb (Table 5).

Features (1–5) are extracted for each role and capture their presence, first POS tag and word, length and position within the roles present for that instance. Features (6–8) further characterize A1. A1-postag is extracted for the following POS tags: DT, JJ, PRP, CD, RB, VB and WP; A1-keyword for the following words: *any*, *anybody*, *anymore*, *anyone*, *anything*, *anytime*, *anywhere*, *certain*, *enough*, *full*, *many*, *much*, *other*, *some*, *specifics*, *too* and *until*. These lists of POS tags and keywords were extracted after manual examination of training examples and aim at signaling whether this role correspond to the focus. Examples of A1 corresponding to the focus and including one of the POS tags or keywords are:

- *[Apparently]*$_{MADV}$, *[the respondents]*$_{A0}$ *do n't think [that an economic slowdown would harm the major investment markets very^{RB} much]*$_{A1}$. (i.e., the responders think it would harm the investements little).

- *[The oil company]$_{A0}$ does n't anticipate [any^{keyword} additional charges]$_{A1}$* (i.e., *the company anticipates no additional charges*).
- *[Money managers and other bond buyers]$_{A0}$ haven't [shown]$_v$ [much^{keyword} interest in the Refcorp bonds]$_{A1}$* (i.e., *they have shown little interest in the bonds*).
- *He concedes H&R Block is well-entrenched and a great company, but says "[it]$_{A1}$ doesn't [grow]$_v$ [fast enough^{keyword} for us]$_{A1}$"* (i.e., it is growing too slow for us).
- [We]$_{A0}$ don't [see]$_v$ [a domestic source for some^{keyword} of our HDTV requirements]$_{A1}$, and that's a source of concern [. . .] (i.e., we see a domestic source for some other of our HDTV requirements)

Features (11–16) correspond to the main verb. VP-words (VP-postag) captures the full sequence of words (POS tags) from the beginning of the VP until the main verb. Features (15–16) check for POS tags as the presence of certain tags usually signal that the verb is not the focus of negation (e.g., *[Thus]$_{MDIS}$, he asserts, [Lloyd's]$_{A0}$ [[ca]$_{MMOD}$n't [react]$_v$ [quickly^{RB}]$_{MMNR}$ [to competition]$_{A1}$]$_{VP}$*).

Features (17–22) tackle the predicate, which includes the main verb and may include other words (typically prepositions). We consider the words in the predicate, as well as the specific thematic roles for each numbered argument. This is useful since PropBank uses different numbered arguments for the same thematic role depending on the frame (e.g., A3 is used as PURPOSE in *authorize.01* and as INSTRUMENT in *avert.01*).

6 Experiments and Results

As a learning algorithm, we use bagging with C4.5 decision trees. This combination is fast to train and test, and typically provides good performance. More features than the ones depicted were tried, but we only report the final set. For example, the parent node for all roles was considered and discarded. We name the model considering all features and trained using bagging with C4.5 trees FOC-DET.

Results over the test split are depicted in Table 6. Simply choosing A1 as the focus yields an accuracy of 42.11. A better baseline is to always pick the last role (58.39 accuracy). Feeding the learning algo-

System	Accuracy
A1	42.11
FIRST	7.00
LAST	58.39
BASIC	61.38
FOC-DET	65.50

Table 6: Accuracies over test split.

rithm exclusively the label corresponding to the last role and flags indicating the presence of roles yields 61.38 accuracy (BASIC baseline).

Having an agreement of 0.72, there is still room for improvement. The full set of features yields 65.50 accuracy. The difference in accuracy between BASIC and FOC-DET (4.12) is statistically significant (Z-value = 1.71). We test the significance of the difference in performance between two systems i and j on a set of ins instances with the Z-score test, where $z = \frac{abs(err_i, err_j)}{\sigma_d}$, err_k is the error made using set k and $\sigma_d = \sqrt{\frac{err_i(1-err_i)}{ins} + \frac{err_j(1-err_j)}{ins}}$.

7 Conclusions

In this paper, we present a novel way to semantically represent negation using focus detection. Implicit positive meaning is identified, giving a thorough interpretation of negated statements.

Due to the lack of corpora annotating the focus of negation, we have added this information to all the negations marked with MNEG in PropBank. A set of features is depicted and a supervised model proposed. The task is highly ambiguous and semantic features have proven helpful.

A verbal negation is interpreted by considering all roles positive except the one corresponding to the focus. This has proven useful as shown in several examples. In some cases, though, it is not easy to obtain the meaning of a negated role.

Consider (Table 3, example 5) *P&G hasn't sold coffee to airlines*. The proposed representation encodes *P&G has sold coffee, but not to airlines*. However, it is not said that the buyers are likely to have been other kinds of companies. Even without fully identifying the buyer, we believe it is of utmost importance to detect that *P&G has sold coffee*. Empirical data (Table 4) shows that over 65% of negations in PropBank carry implicit positive meaning.

References

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 17th international conference on Computational Linguistics*, Montreal, Canada.

Johan Bos and Katja Markert. 2005. Recognising Textual Entailment with Logical Inference. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 628–635, Vancouver, British Columbia, Canada.

Jinho D. Choi, Claire Bonial, and Martha Palmer. 2010. Propbank Instance Annotation Guidelines Using a Dedicated Editor, Jubilee. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Isaac Councill, Ryan McDonald, and Leonid Velikovich. 2010. What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 51–59, Uppsala, Sweden.

David Dowty. 1994. The Role of Negative Polarity and Concord Marking in Natural Language Reasoning. In *Proceedings of Semantics and Linguistics Theory (SALT) 4*, pages 114–144.

Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, pages 1–12, Uppsala, Sweden.

Jaakko Hintikka. 2002. Negation in Logic and in Natural Language. *Linguistics and Philosophy*, 25(5/6).

Laurence R. Horn and Yasuhiko Kato, editors. 2000. *Negation and Polarity - Syntactic and Semantic Perspectives (Oxford Linguistics)*. Oxford University Press, USA.

Laurence R. Horn. 1989. *A Natural History of Negation*. University Of Chicago Press.

Rodney D. Huddleston and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press.

Roser Morante and Walter Daelemans. 2009. Learning the Scope of Hedge Cues in Biomedical Texts. In *Proceedings of the BioNLP 2009 Workshop*, pages 28–36, Boulder, Colorado.

Roser Morante and Caroline Sporleder, editors. 2010. *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*. University of Antwerp, Uppsala, Sweden.

Arzucan Özgür and Dragomir R. Radev. 2009. Detecting Speculations and their Scopes in Scientific Text. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1398–1407, Singapore.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106.

Mats Rooth. 1985. *Association with Focus*. Ph.D. thesis, Univeristy of Massachusetts, Amherst.

Mats Rooth. 1992. A Theory of Focus Interpretation. *Natural Language Semantics*, 1:75–116.

Carolyn P. Rose, Antonio Roque, Dumisizwe Bhembe, and Kurt Vanlehn. 2003. A Hybrid Text Classification Approach for Analysis of Student Essays. In *In Building Educational Applications Using Natural Language Processing*, pages 68–75.

Victor Sánchez Valencia. 1991. *Studies on Natural Logic and Categorial Grammar*. Ph.D. thesis, University of Amsterdam.

Roser Saurí and James Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227–268.

Elly van Munster. 1988. The treatment of Scope and Negation in Rosetta. In *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, Hungary.

Veronika Vincze, Gyorgy Szarvas, Richard Farkas, Gyorgy Mora, and Janos Csirik. 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(Suppl 11):S9+.

Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. 2010. A survey on the role of negation in sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 60–68, Uppsala, Sweden, July.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2009. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis. *Computational Linguistics*, 35(3):399–433.

H. Zeijlstra. 2007. Negation in Natural Language: On the Form and Meaning of Negative Elements. *Language and Linguistics Compass*, 1(5):498–518.