Automatic Generation and Scoring of Positive Interpretations from Negated Statements

Eduardo Blanco and **Zahra Sarabi**
Human Intelligence and Language Technologies Lab
University of North Texas
Denton, TX, 76203
`eduardo.blanco@unt.edu, zahrasarabi@my.unt.edu`

Abstract

This paper presents a methodology to extract positive interpretations from negated statements. First, we automatically generate plausible interpretations using well-known grammar rules and manipulating semantic roles. Second, we score plausible alternatives according to their likelihood. Manual annotations show that the positive interpretations are intuitive to humans, and experimental results show that the scoring task can be automated.

1 Introduction

Negation is an intricate phenomenon present in all human languages (Hoeksema, 2000), and studied from a theoretical perspective since Aristotle. Acquiring and understanding negation is more challenging than language in general: children acquire negation after learning to communicate (Nordmeyer and Frank, 2013), and adults take longer to process negative sentences than positive ones (Clark and Chase, 1972). In any given language, humans communicate in positive terms most of the time, and use negation to express something unusual or an exception (Horn, 1989).

In classical logic, negation is a simple unary operator that reverses the truth value of a proposition. In natural language, negation is always marked (Horn and Wansing, 2015) and it is used to reverse polarity, i.e., turning something affirmative into negative, or something negative into affirmative. Albeit most sentences are affirmative, negation is rather ubiquitous (Morante and Sporleder, 2012): In scientific papers, 13.76% of statements contain a negation (Szarvas et al., 2008); in product reviews, 19% (Councill et al., 2010); and in a selection of Conan Doyle stories, 22.23% (Morante and Daelemans, 2012). In health records, 12.3% of concepts are tagged as negated (Elkin et al., 2005); and in OntoNotes (Hovy et al., 2006), 10.15% of statements contain a verb negated with *not*, *n't* or *never*.

From a theoretical perspective, it is accepted that negation conveys positive meaning (Rooth, 1992; Huddleston and Pullum, 2002). For example, when reading (1) *John doesn't eat meat*, humans intuitively understand that (1a) *John eats something other than meat*, and (1b) *Some people eat meat, but not John*. Extracting positive interpretations from negated statements automatically is not straightforward: a negated statement may convey one or more positive interpretations, and not all positive interpretations are equally likely. For example, from (2) *They didn't order the right parts*, it is very likely that (2a) *They ordered the wrong parts*, but (2b) *Somebody ordered the right parts, but not they* is unlikely.

This paper presents a methodology to automatically extract and score positive interpretations from negated statements, as intuitively done by humans when reading text. A key feature of the work presented here is that it is not tied to any existing approach to extract meaning from text—we generate positive interpretations in plain text, and these positive interpretations can be semantically represented with any existing approach. The main contributions are: (1) procedure to automatically generate plausible positive interpretations from negated statements, (2) annotations scoring plausible positive interpretations,[1] and (3) experiments detailing results with several combinations of features, as well as gold-standard and predicted linguistic information.

---

[1]Available at `http://hilt.cse.unt.edu/`

2 Background and Definitions

Negation is well-understood in grammars, the valid ways to form a negation are well-documented (Quirk et al., 2000; van der Wouden, 1997). Negation can be expressed by verbs (e.g., avoid doing any look-up), nouns (e.g., the absence of any phonic sequence), adjectives (e.g., it is pointless to argue with a fool), adverbs (e.g., I never tried Persian food before), prepositions (e.g., you can always exchange it without a problem), determiners (e.g., the new law has no direct implications to international shipping), pronouns (e.g., nobody will keep election promises).

Huddleston and Pullum (2002) distinguish four negation types:

- Verbal if the marker of negation is grammatically associated with a verb, e.g., *I did not see anything*, non-verbal if it is associated with a dependent of the verb, e.g., *I saw nothing*.
- Synthetic if the negation mark has a function besides marking a negation, e.g., *[Nobody]*_{AGENT} *liked it*, analytic otherwise, e.g., *Not many people liked it*.
- Clausal if the negation yields a negative clause, e.g., *The terms aren't negotiable*, subclausal otherwise, e.g., *The terms are non-negotiable*.
- Ordinary if the negation indicates that something is not the case, e.g., *That car does not drive smooth*, metalinguistic if it does not dispute the truth but rather reformulates a statement, e.g., *That TV is not small, it is tiny*.

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

2.1 Positive Interpretations.

In philosophy and linguistics, it is generally accepted that negation conveys positive meanings (Horn, 1989). These positive meanings range from implicatures, i.e., what is suggested in an utterance even though neither expressed nor strictly implied (Blackburn, 2008), to entailments. Other terms used in the literature include implied meanings (Mitkov, 2005), implied alternatives (Rooth, 1985) and semantically similars (Agirre et al., 2013). We do not strictly fit into any of this terminology, we reveal positive interpretations as intuitively done by humans when reading text.

2.2 Scope and Focus.

From a theoretical perspective, it is accepted that negation has scope and focus, and that the focus—not just the scope—yields positive interpretations (Horn, 1989; Rooth, 1992; Taglicht, 1984). Scope is "the part of the meaning that is negated" and focus "the part of the scope that is most prominently or explicitly negated" (Huddleston and Pullum, 2002).

Consider the following statement in the context of the recent refuge crisis: (3) *Mr. Haile was not looking for heaven in Europe*. By definition, scope refers to "all elements whose individual falsity would make the negated statement strictly true", and focus is "the element of the scope that is intended to be interpreted as false to make the overall negative true" (Huddleston and Pullum, 2002). The falsity of any of the truth conditions below makes statement (3) true, thus the scope of the negation is (3a–3d):

3a. Somebody was looking for something somewhere. [verb *looking*]
3b. Mr. Haile was looking for something somewhere. [AGENT of looking, *Mr. Haile*]
3c. Somebody was looking for heaven somewhere. [THEME of looking, *heaven*]
3d. Somebody was looking for something in Europe. [LOCATION of looking, *in Europe*]

Determining the focus is almost always more challenging than the scope. The challenge lies on determining which of the truth conditions (3a–3d) is intended to be interpreted as false to make the negated statement true: all of them qualify, but some are more likely. A natural reading of statement (3) suggests that *Mr. Haile was looking for something (a regular life, a job, etc.) in Europe, but not heaven*. Determining that the focus is *heaven*, i.e., that everything in statement (3) is actually positive except the THEME of *looking*, is the key to reveal the intended positive interpretation. It is worth noting that other foci yield unlikely interpretations, e.g., *Somebody was looking for heaven in Europe, but not Mr. Haile* (3b, AGENT), *Mr. Haile was looking for heaven somewhere, but not in Europe* (3d, LOCATION). Note that (1) scope on its own does not yield positive interpretations, and (2) some negated statements convey several likely positive interpretations, e.g., statement (1) in Section 5, Table 3.

3 Previous Work

Within computational linguistics, approaches to process negation are shallow, or target scope and focus detection. Popular semantic representations such as semantic roles (Palmer et al., 2005; Baker et al., 1998) or AMR (Banarescu et al., 2013) do not reveal the positive interpretations we target in this paper. Shallow approaches are usually application-specific. In sentiment and opinion analysis, negation has been reduced to marking as negated all words between a negation cue and the first punctuation mark (Pang et al., 2002), or within a five-word window of a negation cue (Hu and Liu, 2004). The examples throughout this paper show that these techniques are insufficient to reveal implicit positive interpretations.

3.1 Scope Annotation and Detection

Scope of negation detection has received a lot of attention, mostly using two corpora: BioScope in the medical domain (Szarvas et al., 2008) and CD-SCO (Morante and Daelemans, 2012). BioScope annotates negation cues and linguistic scopes exclusively in biomedical texts. CD-SCO annotates negation cues, scopes, and negated events or properties in selected Conan Doyle stories.

There have been several supervised proposals to detect the scope of negation using BioScope and CD-SCO (Özgür and Radev, 2009; Øvrelid et al., 2010). Automatic approaches are mature (Abu-Jbara and Radev, 2012): F-scores are 0.96 for negation cue detection, and 0.89 for negation cue and scope detection (Velldal et al., 2012; Li et al., 2010). Outside BioScope and CD-SCO, Reitan et al. (2015) present a negation scope detector for tweets, and show that it improves sentiment analysis. As shown in Section 2, scope detection is insufficient to reveal positive interpretations from negated statements.

3.2 Focus Annotation and Detection

While focus of negation has been studied for decades in philosophy and linguistics (Section 2), corpora and automated tools are scarce. Blanco and Moldovan (2011) annotate focus of negation in the 3,993 negations marked with ARGM-NEG semantic role in PropBank (Palmer et al., 2005). Their annotations, PB-FOC, were used in the *SEM-2012 Shared Task (Morante and Blanco, 2012). Their guidelines require annotators to choose as focus the semantic role that "is most prominently negated" or the verb. If several roles may be the focus, they prioritize "the one that yields the most meaningful implicit [positive] information", but do not specify what *most meaningful* means. Consider again statement (1) *John doesn't eat meat*. Their approach would determine that the focus is the THEME of *eat*, *meat*, because it arguably yields the "most meaningful implicit [positive] information" (using our terminology, positive interpretation): *John eats something other than meat*. By design, they ignore other valid positive interpretations, e.g., *Some people eat meat, but not John*. In this paper, we improve upon their work: instead of extracting the "most meaningful" positive interpretation from a negated statement, we generate several positive interpretations and score them according to their likelihood.

Anand and Martell (2012) present a complimentary approach to annotate focus of negation. They refine PB-FOC and argue that positive interpretations arising from scalar implicatures and negraising predicates should be separated from those arising from focus detection. According to their annotations, 27.4% of negations with a focus annotated in PB-FOC do not actually have a focus. Blanco and Moldovan (2012) introduce the concept of fine-grained foci and refine the annotations in PB-FOC by annotating foci at the token level, and Matsuyoshi et al. (2014) annotate focus of negation in Japanese. In this paper, we are not concerned about annotating focus of negation per se, but about extracting positive interpretations from negated statements as intuitively understood by humans.

Automatic systems to detect the focus of negation (and reveal up to one positive interpretation) in English texts are trained using PB-FOC. Blanco and Moldovan (2011) obtain an accuracy of 65.5 using supervised machine learning and features derived from gold-standard linguistic information, and Blanco and Moldovan (2014) report an F-measure of 64.1. Rosenberg and Bergler (2012) report an F-measure of 58.4 using 4 linguistically sound heuristics and predicted linguistic information, and Zou et al. (2014) an F-measure of 65.62 using contextual discourse information. Unlike the work presented here, none of these systems attempts to extract and rank several positive interpretations from one negated statement.

Neg. statement		In 1995, Murdoch bought [the rest of Star TV]$_{ARG_1}$ that [he]$_{ARG_0}$ did [not]$_{ARGM\text{-}NEG}$ [own]$_{verb}$
Positive counterpart	after Step 1	In 1995, Murdoch bought the rest of Star TV that he did own
	after Step 2	In 1995, Murdoch bought the rest of Star TV that he owned
	after Step 3	In 1995, Murdoch bought the rest of Star TV that he owned (no change)
Plausible positive interpretations	from ARG_0	In 1995, Murdoch bought the rest of Star TV that [somebody] owned, but not *he*
	from ARG_1	In 1995, Murdoch bought [something] that he owned, but not *the rest of Star TV*
	from *verb*	In 1995, Murdoch bought the rest of Star TV that he [some verb], but not *owned*

Table 1: Negated statement, and automatically generated positive counterpart and plausible positive interpretations.

4 Corpus Creation

Our goal is to create a corpus of negated statements and their positive interpretations as intuitively understood by humans. We put a strong emphasis on automation. First, given a negated statement, we automatically generate plausible positive interpretations following a battery of linguistically motivated deterministic rules (Section 4.1). Second, we collect manual annotation to score the plausible positive interpretations according to their likelihood (Section 4.2). We then use these manually obtained scores to learn models that automatically score positive interpretations (Section 6).

We decided to work on top of OntoNotes (Hovy et al., 2006) instead of plain text or other corpora for several reasons. First, OntoNotes includes gold linguistic annotations such as part-of-speech tags, parse trees and semantic roles. Second, state-of-the-art role labelers trained with Propbank achieve F-measures of 0.835 (Lewis et al., 2015), and we use semantic roles to generate positive interpretations. Third, unlike BioScope, CD-SCO and PB-FOC (Section 2), OntoNotes includes sentences from several genres, e.g., newswire, broadcast news and conversations, magazines, the web.

4.1 Generating Positive Interpretations

OntoNotes[2] is a large corpus containing 63,918 sentences. Annotating all positive interpretations from all negations is outside of the scope of this paper. Instead, we target selected representative negations.
Selecting Negated Statements. We first selected all verbs negated with ARGM-NEG semantic role and obtained 6,617 verbal negations. After examining the negated verbs, it became clear that negation is not uniformly distributed across verbs in OntoNotes,

it roughly follows Zipf's law. In order to alleviate the annotation effort while accounting for all negated verbs in OntoNotes, we randomly selected up to 5 negations for each verb. The number of negated statements selected is 600.
Converting Negated Statements into Their Positive Counterparts. We apply 3 steps inspired after the grammatical rules to form negation detailed by Huddleston and Pullum (2002, Ch. 9):

1. Remove the negation mark by removing the tokens within ARGM-NEG semantic role.
2. Remove auxiliaries, expand contractions, and fix third-person singular and past tense. For example (before: after), *doesn't go*: *goes*, *didn't go*: *went*, *won't go*: *will go*, We use a standard list of irregular verbs,[3] and grammar rules to convert to third-person singular and past tense based on orthographic patterns.
3. Rewrite negatively-oriented polarity-sensitive items. For example (before: after), *anyone*: *someone*, *any longer*: *still*, *yet*: *already*. *at all*: *somewhat*. We use the correspondences between negatively-oriented and positively-oriented polarity-sensitive items by Huddleston and Pullum (2002, pp. 831).

Generating Positive Interpretations. Once the positive counterpart is obtained, we generate positive interpretations by rewriting each semantic role or the (originally negated) verb. Among others, we use the following rewriting rules: ARG_0–ARG_4: someone / some people / something, ARGM-TMP: at some point of time, ARGM-LOC: somewhere, ARGM-MNR: in some manner, ARGM-CAU: because of something and ARGM-PRP: to do something. Additionally, if the semantic role starts with a preposition, we also include it, e.g., *gave [to John]*$_{ARG_2}$:

---

gave to someone, but not John. This methodology generated 1,888 positive interpretations from the 600 selected negations (average: 3.15).

Table 1 exemplifies the 3 steps to transform a negated statement into its positive counterpart, and the positive interpretations generated. Additional examples are provided in Table 3.

We acknowledge that some of the positive interpretations we generate automatically are not as specific or intuitive as carefully crafted, manually generated interpretations could be. For example, from *[John]*$_{\text{ARG}_0}$ *does[n't]*$_{\text{ARGM-NEG}}$ *[know]*$_{verb}$ *[the details about how they met]*$_{\text{ARG}_1}$, the proposed methodology would generate, among others, *John knows about something, but not about the details of how they met*. A better interpretations that we do not generate is *John knows something about how they met, but not the details*. We argue that generating interpretations automatically is the only option in order to incorporate this work into an NLP pipeline, and reserve for future work generating positive interpretations beyond rewriting semantic roles.

4.2 Ranking Positive Interpretations

Once positive interpretations were automatically generated, we asked annotators to rank them. Annotators were presented with one negated statement and one positive interpretation at a time, and were asked *Given the negated statement above, do you think the statement [positive interpretation] below is true?* They only had access to the text in the original negated statement, the positive interpretation, and the previous and next sentences as context. We did not display semantic role information for the original negated statement, its positive counterpart or the semantic role from which the positive interpretation was generated. As we shall see (Section 5.1), context often helps scoring interpretations.

Annotators were required to answer with a score from 0 to 5, were 0 means *absolutely disagree* and 5 means *absolutely agree*. We did not provide descriptions for intermediate scores or used additional categorical labels. This simple guidelines were sufficient to reliably score plausible positive interpretations automatically generated.

Sem. role	#	% sent	Mean	SD
ARG_0	392	65.17%	4.07	1.03
ARG_1	520	86.00%	4.55	0.98
ARG_2	95	15.83%	4.40	1.31
ARG_3	2	0.33%	4.50	0.50
ARG_4	5	0.83%	5.00	0.00
ARGM-ADV	101	15.67%	2.41	1.72
ARGM-CAU	13	2.17%	2.46	1.91
ARGM-DIR	11	1.83%	2.00	2.30
ARGM-EXT	12	2.00%	3.75	1.53
ARGM-LOC	21	3.33%	3.76	1.69
ARGM-MNR	37	6.00%	4.54	1.03
ARGM-PRP	7	1.16%	4.43	0.71
ARGM-TMP	81	12.50%	3.99	1.52
Verb	600	100.00%	2.17	1.41
All	1,888	100.00%	3.52	1.63

Table 2: Basic corpus analysis. For each semantic role, we show the number of positive interpretations generated (#), the percentage of sentences for which a positive interpretation is generated (% sent), and the mean and standard deviation (SD) of the annotated scores.

5 Corpus Analysis

On average, we generated 3.15 positive interpretations per negation (standard deviation: 0.82), and 74% of negations have at least one interpretation scored 4 or higher. Basic counts and statistics for the annotations are provided in Table 2. Overall, we annotated 1,888 positive interpretations generated from 600 sentences, or equivalently, from 600 verbs negated with ARGM-NEG semantic role in OntoNotes. Overall mean score is 3.52 (out of 5) and overall standard deviation, 1.63. The 25th percentile is 2.0, the 50th percentile is 4.0 and the 75th percentile is 5.0. These numbers show that most positive interpretations automatically generated are deemed likely by annotators, and over 25% are scored with a 5 (out of 5).

In general, positive interpretations generated from numbered roles (ARG_0–ARG_4) are scored higher than the ones generated from modifiers (ARGM-ADV, ARGM-CAU, ARGM-DIR, etc.). Also, positive interpretations generated from infrequent roles are generally ranked higher, e.g., ARG_4 and ARG_3 vs. ARG_0, ARGM-PRP and ARGM-MNR vs. ARGM-ADV.

Annotation Quality. In order to ensure annotation quality, we calculated inter-annotator Pearson correlation. Kappa and other agreement measures de-

	Negated statement, context if relevant to determining scores, and all positive interpretations	Score
1	Context, previous statement: *That change will obviously impact third and fourth quarter earnings for the industry in general, he added.* Negated statement: *[He]$_{\text{ARG}_0}$ did[n't]$_{\text{ARGM-NEG}}$ [forecast]$_{verb}$ [Phillips' results]$_{\text{ARG}_1}$.* Context, next statement: *But security analysts say Phillips will be among the companies hard-hit by weak chemical prices and will probably post a drop in third-quarter earnings.*	
	- Interpretation 1.1 [ARG$_0$]: Somebody forecasted Phillips' results, but not *he*.	5
	- Interpretation 1.2 [ARG$_1$]: He forecasted something, but not *Phillips' results*.	5
2	Negated statement: *In 1995, Murdoch bought [the rest of Star TV]$_{\text{ARG}_1}$ [he]$_{\text{ARG}_0}$ did [not]$_{\text{ARGM-NEG}}$ [own]$_{verb}$.*	
	- Interpretation 2.1 (ARG$_0$): In 1995, Murdoch bought the rest of Star TV somebody owned, but not *he*.	5
	- Interpretation 2.2 (ARG$_1$): In 1995, Murdoch bought something he owned, but not *the rest of Star TV*.	0
3	Negated statement: *[You]$_{\text{ARG}_0}$ [can]$_{\text{ARGM-MOD}}$[not]$_{\text{ARGM-NEG}}$ [run]$_{verb}$ [a country with 23 million people]$_{\text{ARG}_1}$ [with revenues of 16 billion to 20 billion dollars]$_{\text{ARGM-MNR}}$.*	
	- Interpretation 3.1 (ARG$_0$): Somebody can run a country with 23 million people with revenues of 16 billion to 20 billion dollars, but not *You*.	1
	- Interpretation 3.21 (ARG$_1$): You can run something with revenues of 16 billion to 20 billion dollars, but not *a country with 23 million people*.	5
	- Interpretation 3.3 (ARGM-MNR): You can run a country with 23 million people in some manner, but not with *with revenues of 16 billion to 20 billion dollars*.	5
4	Negated statement: *Do [not]$_{\text{ARGM-NEG}}$ [utter]$_{verb}$ [a word]$_{\text{ARG}_1}$.*	
	- Interpretation 4.1 (ARG$_1$): Utter something, but not *a word*.	0

Table 3: Annotation examples. We show all positive interpretations automatically generated and their scores (out of 5).

signed for categorical labels are not well-suited for our annotation task, since not all disagreements between numeric scores are the same, e.g., 4 vs. 5 should be counted as relatively high agreement, and 1 vs. 5 should be counted as high disagreement. Overall Pearson correlation is 0.761.

5.1 Annotation Examples

Table 3 presents annotation examples. We show the original negated statement including semantic role annotations from OntoNotes (square brackets), all positive interpretations automatically generated, and their scores. We also include context (previous and next sentence) if it helps determining scores.

Example (1) shows that context sometimes is vital to scoring plausible positive interpretations. Given *He didn't forecast Phillips' results* in isolation, it is uncertain if he forecasted anything at all, or whether somebody forecasted Phillips' results. However, the previous statement makes certain (5/5) the interpretation generated from ARG$_1$: he forecasted earnings for the industry in general. Similarly, the next statement makes very likely (5/5) the interpretation generated from ARG$_0$: other people (security analysts) made forecasts about Phillips. In this example, 2

positive interpretations are generated from one negation, and they are assigned the highest score (5/5).

The positive interpretations generated from example (2) can be annotated without context. Somebody other than Murdoch had to own the rest of Star TV that he bought in 1995 (score 5/5), and people cannot buy what they already own (score 0/5).

Example (3) presents a positive interpretation generated from ARG$_0$ that is scored low (1/5); recall that the mean score for interpretations generated from ARG$_0$ is 4.07 and the standard deviation 1.03 (Table 2). In this negated statement, the indefinite *you* refers to an unspecified person, thus it is not the case that *somebody can run a country with 23 million people with revenues of 16 billion to 20 billion dollars* (Interpretation 3.1, score 1/5). Interpretations 3.2 and 3.3, however, are scored high: annotators correctly understood that given the negated statement, it is the case that *You can run something with revenues of 16 billion to 20 billion dollars, but not a country with 23 million people* (you can run a country with less people with that revenue), and *You can run a country with 23 million people in some manner, but not with revenues of 16 billion to 20 billion dollar* (you could run it with more revenue).

Finally, example (4) presents a short statement from which only one positive interpretations is generated. Annotators were asked whether given *Do not utter a word*, they think *Utter something, but not a word* is true. They correctly annotated that this positive interpretations is invalid (score 0/5). Indeed, the negated statement in example (4) can only be interpreted as an order to not utter anything.

6 Learning to Score Positive Interpretations

We follow a standard supervised machine learning approach. The 1,888 positive interpretations along with their scores become instances, and we divide them into training (80%) and test splits (20%) making sure that all interpretations generated from a sentence are assigned to either the training or test splits. Note that splitting instances randomly would not be sound: training with some interpretations generated from a negated statement, and testing with the rest of interpretations generated from the same statement would be an unfair evaluation.

We trained a Support Vector Machine (SVM) for regression with RBF kernel using scikit-learn (Pedregosa et al., 2011), which in turn uses LIBSVM (Chang and Lin, 2011). The feature set and SVM parameters (C and γ) were tuned using 10-fold cross-validation with the training set, and results were calculated using the test set.

6.1 Feature Selection

We tried features derived exclusively from the negated statement from which the positive interpretation was generated, more specifically, we extract features from the negated *verb* or semantic role (*sem_role*) used to generate the positive interpretation, from both of them (*verb-sem_role*) and from the verb-argument structure of *verb* (*verbarg-struct*), i.e., all semantic roles of *verb*.

Verb features are straightforward and account for the verb word form and part-of-speech tag.

Sem_role features include the label of the semantic role from which the positive interpretation was generated (*sem_role_label*), its length (number of tokens), and the word form and part-of-speech tag of its head. Additionally, we add standard features in semantic role labeling (Gildea and Jurafsky, 2002):

the syntactic nodes (NP, PP, etc.) of the semantic role and its parent in the parse tree, as well as the left and right siblings, if any.

Verb-sem_role features are also standard in role labeling. We include a flag indicating whether the verb occurs before or after the semantic role in the negated statement (not in the positive counterpart), syntactic node of the lowest common ancestor between *verb* and *sem_role*, and the syntactic path.

Finally, *verbarg-struct* features encode characteristics of the verb-argument structure to which *verb* and *sem_role* belong to. Namely, we added flags indicating semantic role presence, and features indicating the first and last semantic roles in order of appearance in the negated statement. We also included the syntactic nodes of semantic roles and their heads.

We exemplify all features with Interpretation 3.1 generated from Statement 3 in Table 3:

- *Verb* features: verb_wf=run, verb_pos=VBP.
- *Sem_role* features: label=ARG_0, num_tokens=1, head_wf=You, head_pos=PRP, synt_node=NP, synt_node_parent=S, synt_node_left=NONE, synt_node_right=VP
- *Verb-sem_role* features: direction=after, lowest_ancestor=S, synt_path=VP+VP+S-NP
- *Verbarg-struct* features: ARG_0=1, ARGM-MOD=1, ARG_1=1, ARGM-MNR=1, first=ARG_0, last=ARGM-MNR, ARG_1_head=country, ARG_1_pos=NN, etc.

7 Experimental Results

We report results obtained with several combinations of features in Table 5. We detail results obtained with features extracted from gold-standard and predicted linguistic annotations (part-of-speech tags, parse trees, semantic roles, etc.) as annotated in the *gold* and *auto* files from the CoNLL-2011 Shared Task release of OntoNotes (Pradhan et al., 2011). All models are trained with gold-standard linguistic annotations, and tested with either gold-standard or predicted linguistic annotations.

Testing with gold-standard linguistic annotations. Using only the label of the semantic role from which the positive interpretation was generated (*sem_role_label*), yields a Pearson correlation of 0.603. Using *Verb* features is virtually useless (Pearson correlation: -0.025), this is due to the

Source	Feature description
Verb	Word form and part-of-speech tag of *verb*
sem_role	Semantic role label of sem_role, *sem_role_label*
	Number of tokens in *sem_role*
	Word form and part-of-speech tag of head of *sem_role*
	Syntactic node of *sem_role*, its parent and left and right siblings in the parse tree
verb-sem_role	Whether *verb* occurs before or after than *sem_role* in the negated statement
	Syntactic node of lowest common ancestor of *verb* and *sem_role*
	Syntactic paths from *verb* to *sem_role*
verbarg-struct	Flags indicating whether *verb* has each possible semantic role
	Semantic role labels of the first and last roles of *verb*
	Syntactic nodes and heads of each semantic role attaching to *verb*

Table 4: Features used to score positive interpretations from negated statements. Features are extracted from the negated *verb* or semantic role (*sem_role*) from which the positive interpretation was generated, from both of them (*verb-sem_role*), or from the verb-argument structure of *verb* (verbarg-struct), i.e., from all semantic roles of *verb*.

	Gold	Predicted
sem_role_label feature	0.603	0.642
verb features	-0.025	-0.022
verb + sem_role features	0.630	0.648
verb + sem_role + verb-sem_role features	0.627	0.638
verb + sem_role + verb-sem_role + verbarg-struct features	0.642	0.650

Table 5: Pearson correlations in the test set using gold-standard and predicted linguistic annotations (part-of-speech tags, parse trees and semantic roles). Results are provided using sem_role_label as only feature, and using several features incrementally. Number of test instances with gold-standard linguistic annotations is 378, and with predicted annotations, 268.

fact that our corpus includes at most 5 instances for each negated verb in OntoNotes (Section 4). Adding *sem_role* features yields a correlation of 0.630, and incorporating *verb-sem_role* features is useless (Pearson: 0.627). Considering all features, however, yields the highest correlation, 0.642.

Testing with predicted linguistic annotations. We assigned 20% of instances to the test split, totalling 378 instances (Section 6). However, some of these instances cannot be obtained using predicted role labels: a missing or incorrect semantic role will unequivocally lead to positive interpretations that are not in our corpus and thus evaluation is not straightforward. Results presented with predicted linguistic annotations are calculated using only the 268 (out of 378) positive interpretations that are generated from predicted semantic roles and are also generated (and thus annotated in our corpus) from gold-standard linguistic annotations.

Results using only *sem_role_label* feature (0.642) are better or very similar than using any combination of features (0.638–0.650) except *Verb* features alone, which perform poorly as explained earlier.

These results should be taken with a grain of salt: the number of test instances is much lower (268 vs. 378). Additionally, these 268 test instances correspond to positive interpretations generated from semantic roles that were predicted correctly automatically. Role labels are predicted better for shorter sentences without complicated syntactic structure; positive interpretations for this kind sentences are also easier to score, e.g., Statement 4 in Table 3.

8 Conclusions

Humans intuitively understand negated statements in positive terms when reading text. This paper presents an automated methodology to generate plausible positive interpretations from verbal negation, and score them based on their likelihood. We use simple grammar rules and manipulate semantic roles to generate positive interpretations. Experimental results show that these interpretations can be scored automatically using standard supervised machine learning techniques.

An annotation effort shows that most positive interpretations automatically generated are likely

(scores ≥ 4), thus the amount of positive meaning revealed by the methodology presented here is substantial (on average, 3.15 interpretations are generated per negation). We believe that more annotations and a learning algorithm that scores jointly all positive interpretations generated from each negation (as opposed to individually) would yield better results.

References

Amjad Abu-Jbara and Dragomir Radev. 2012. Umichigan: A conditional random field model for resolving the scope of negation. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 328–334. Association for Computational Linguistics.

Eneko Agirre, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, and Weiwei Guo. 2013. *sem 2013 shared task: Semantic textual similarity. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity*, pages 32–43, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

Pranav Anand and Craig Martell. 2012. Annotating the focus of negation in terms of questions under discussion. In *Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics*, ExProM '12, pages 65–69, Stroudsburg, PA, USA. Association for Computational Linguistics.

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.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria, August. Association for Computational Linguistics.

Simon Blackburn. 2008. *The Oxford Dictionary of Philosophy*. Oxford University Press.

Eduardo Blanco and Dan Moldovan. 2011. Semantic representation of negation using focus detection. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 581–589, Portland, Oregon, USA, June. Association for Computational Linguistics.

Eduardo Blanco and Dan Moldovan. 2012. Fine-grained focus for pinpointing positive implicit meaning from negated statements. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 456–465, Montréal, Canada, June. Association for Computational Linguistics.

Eduardo Blanco and Dan Moldovan. 2014. Retrieving implicit positive meaning from negated statements. *Natural Language Engineering*, 20:501–535, 10.

Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May.

H. H. Clark and W. G. Chase. 1972. On the process of comparing sentences against pictures. *Cognitive Psychology*, 3(3):472–517, July.

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, July. University of Antwerp.

Peter L. Elkin, Steven H. Brown, Brent A. Bauer, Casey S. Husser, William Carruth, Larry R. Bergstrom, and Dietlind L. Wahner-Roedler. 2005. A controlled trial of automated classification of negation from clinical notes. *BMC Medical Informatics and Decision Making*, 5(13).

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Comput. Linguist.*, 28(3):245–288, September.

Jack Hoeksema. 2000. Negative polarity items: triggering, scope,and c-command. In L. Horn and Y. Kato, editors, *Negation and polarity*, chapter 4, pages 115–146. Oxford University Press, Oxford.

Laurence R. Horn and Heinrich Wansing. 2015. Negation. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2015 edition.

Laurence R. Horn. 1989. *A natural history of negation*. Chicago University Press, Chicago.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: the 90% Solution. In *NAACL '06: Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers on XX*, pages 57–60, Morristown, NJ, USA. Association for Computational Linguistics.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference*

on Knowledge discovery and data mining, pages 168–177, New York, NY, USA. ACM.

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

Mike Lewis, Luheng He, and Luke Zettlemoyer. 2015. Joint a* ccg parsing and semantic role labelling. In *Empirical Methods in Natural Language Processing*.

Junhui Li, Guodong Zhou, Hongling Wang, and Qiaoming Zhu. 2010. Learning the Scope of Negation via Shallow Semantic Parsing. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 671–679, Beijing, China, August. Coling 2010 Organizing Committee.

Suguru Matsuyoshi, Ryo Otsuki, and Fumiyo Fukumoto. 2014. Annotating the focus of negation in japanese text. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1743–1750, Reykjavik, Iceland, May. European Language Resources Association (ELRA). ACL Anthology Identifier: L14-1606.

Ruslan Mitkov. 2005. *The Oxford handbook of computational linguistics*. Oxford University Press.

Roser Morante and Eduardo Blanco. 2012. *SEM 2012 Shared Task: Resolving the Scope and Focus of Negation. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM 2012)*, pages 265–274, Montréal, Canada, June.

Roser Morante and Walter Daelemans. 2012. Conandoyle-neg: Annotation of negation in conan doyle stories. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, Istanbul*.

Roser Morante and Caroline Sporleder. 2012. Modality and negation: An introduction to the special issue. *Comput. Linguist.*, 38(2):223–260, June.

Ann E Nordmeyer and Michael C Frank. 2013. Measuring the comprehension of negation in 2-to 4-year-old children. *Proceedings of the 35th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society*.

Lilja Øvrelid, Erik Velldal, and Stephan Oepen. 2010. Syntactic Scope Resolution in Uncertainty Analysis. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1379–1387, Beijing, China, August. Coling 2010 Organizing Committee.

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, August. Association for Computational Linguistics.

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

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pages 79–86. Association for Computational Linguistics, July.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Sameer Pradhan, Lance Ramshaw, Mitchell Marcus, Martha Palmer, Ralph Weischedel, and Nianwen Xue. 2011. Conll-2011 shared task: Modeling unrestricted coreference in ontonotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–27, Portland, Oregon, USA, June. Association for Computational Linguistics.

Randolph Quirk, Sidney Greenbaum, and Geoffrey Leech. 2000. *A comprehensive grammar of the English language*. Longman, London.

Johan Reitan, Jørgen Faret, Björn Gambäck, and Lars Bungum. 2015. Negation scope detection for twitter sentiment analysis. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 99–108, Lisboa, Portugal, September. Association for Computational Linguistics.

Mats Rooth. 1985. *Association with focus*. Ph.D. thesis, Dept. of Linguistics, University of Massachusetts, Amherst.

Mats Rooth. 1992. A theory of focus interpretation. *Natural language semantics*, 1(1):75–116.

Sabine Rosenberg and Sabine Bergler. 2012. Uconcordia: Clac negation focus detection at *sem 2012. In *SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 294–300, Montréal, Canada, 7-8 June. Association for Computational Linguistics.

György Szarvas, Veronika Vincze, Richárd Farkas, and János Csirik. 2008. The BioScope corpus: annotation

for negation, uncertainty and their scopein biomedical texts. In *Proceedings of BioNLP 2008*, pages 38–45, Columbus, Ohio, USA. ACL.

Josef Taglicht. 1984. *Message and emphasis: On focus and scope in English*, volume 15. Addison-Wesley Longman Limited.

Ton van der Wouden. 1997. *Negative contexts: collocation, polarity, and multiple negation*. Routledge, London.

Erik Velldal, Lilja Ovrelid, Jonathon Read, and Stephan Oepen. 2012. Speculation and negation: Rules, rankers, and the role of syntax. *Comput. Linguist.*, 38(2):369–410, June.

Bowei Zou, Guodong Zhou, and Qiaoming Zhu. 2014. Negation focus identification with contextual discourse information. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 522–530, Baltimore, Maryland, June. Association for Computational Linguistics.