

A Corpus of Negations and their Underlying Positive Interpretations

Zahra Sarabi, Erin Killian, Eduardo Blanco and Alexis Palmer

University of North Texas

zahrasarabi@my.unt.edu, erinkillian@my.unt.edu,

eduardo.blanco@unt.edu, alexis.palmer@unt.edu

Abstract

Negation often conveys implicit positive meaning. In this paper, we present a corpus of negations and their underlying positive interpretations. We work with negations from Simple Wikipedia, automatically generate potential positive interpretations, and then collect manual annotations that effectively rewrite the negation in positive terms. This procedure yields positive interpretations for approximately 77% of negations, and the final corpus includes over 5,700 negations and over 5,900 positive interpretations. We also present baseline results using seq2seq neural models.

1 Introduction

Negation is present in every human language. It is in the first place a phenomenon of semantical opposition. As such, negation relates an expression e to another expression with a meaning that is in some way opposed to the meaning of e (Horn and Wansing, 2015). Sentences containing negation are generally (a) less informative than affirmative ones (e.g., *Milan is not the capital of Italy* vs. *Rome is the capital of Italy*), (b) morphosyntactically more marked—all languages have negative markers while only a few have affirmative markers, and (c) psychologically more complex and harder to process (Horn and Wansing, 2015).

Negation often conveys implicit positive meanings (Rooth, 1992). This meaning ranges from implicatures to entailments, and we refer to it as *positive interpretations*. Consider the following text from Simple Wikipedia:¹ *An abjad is an alphabet in which all its letters are consonants. Though vowels can be added in some abjads, they are not needed to write a word correctly. Some examples of abjads are the Arabic alphabet and the*

¹<https://simple.wikipedia.org/wiki/Abjad>

1	Mr. Smith apologized for <u>not</u> getting involved. <i>Mr. Smith apologized for staying passive.</i>
2	I <u>never</u> heard of this guy before they started doing these commercials on television and radio. <i>I heard of this guy after they started doing these commercials on television and radio.</i>
3	In Hinduism, beef is <u>not</u> allowed to be eaten. <i>In Hinduism, chicken is allowed to be eaten. In other religions, beef is allowed to be eaten.</i>

Table 1: Three sentences containing a negation and their positive interpretations (italics).

Hebrew alphabet. Humans intuitively understand that the negation (second sentence) implies the following positive interpretation: *Though vowels can be added in some abjads, only consonants are needed to write a word correctly*. Table 1 shows three sentences containing negation and their underlying positive interpretations. Positive interpretations do not have any negation cues (e.g., *not*, *never*) and Example 3 shows that some negations may have more than one underlying positive interpretation depending on the context.

Revealing the underlying positive interpretation of negation is challenging. First, we need to identify which tokens are intended to be negated (e.g., *getting involved* and *before* in Examples 1 and 2 from Table 1). Second, we need to rewrite those tokens to generate an actual positive interpretation (e.g., *getting involved: staying passive*).

This paper presents a corpus of negations and their underlying positive interpretations.² The main contributions are: (a) deterministic procedure to generate potential positive interpretations from negations, (b) corpus of negations and their positive interpretations manually annotated, (c) detailed analysis including which subtrees in the dependency tree are more likely to be rewritten and qualitative analysis of positive interpreta-

²Available at: <https://zahrasarabi.com>

tions. Additionally, we establish baseline results with sequence-to-sequence neural models.

2 Background and Definitions

Negation is well-understood in grammars and the valid ways to express negation are documented (Quirk et al., 2000; van der Wouden, 1997). In this paper, we focus on verbal negations, i.e., when the negation mark—usually an adverb such as *never* and *not*—is grammatically associated with a verb.

Positive Interpretations. In philosophy and linguistics, it is accepted that negation conveys positive meaning (Horn, 1989). This positive meaning ranges 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 similar (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. Note that a *positive* interpretation is a statement that does not contain negation, not a statement that conveys positive sentiment. For example, *The seller didn't ship the right parts* implicitly conveys *The seller shipped the wrong parts*, which has negative sentiment.

Potential Positive Interpretations. Given a sentence containing negation, we use the term *potential positive interpretation* to refer to positive interpretations that are automatically generated by replacing selected tokens with a placeholder. If the placeholder can be rewritten so that the result is an affirmative statement that is true given the original sentence, *potential* positive interpretations become *actual* positive interpretations.

Negation and natural language understanding. Generating positive interpretations from negation has several potential applications.

First, while neural machine translation is in general superior to phrase-based methods, that is not the case when translating negation (Bentivogli et al., 2016). Since our positive interpretations effectively rewrite negation-containing sentences to remove the negation, we argue that they have the potential to help machine translation.

Second, current benchmarks for natural language inference (Bowman et al., 2015), do not include challenging examples with negation. As a result, state-of-the-art approaches (Chen et al.,

2017) trained on these benchmarks are unable to solve text-hypothesis pairs that contain negation. Indeed, we tested the aforementioned systems with 100 text-hypothesis pairs from our corpus (text: sentence with negation, hypothesis: positive interpretation with correctness score of 4; see examples in Table 7), and discovered that 48 of them are predicted *contradiction*, 30 *neutral* and only 22 *entailment* (the correct prediction is entailment for all of them). While relatively small, we argue that the corpus presented here is a step towards language understanding when negation is present.

3 Previous Work

From a theoretical perspective, it is accepted that negation has scope and focus, and that the focus yields positive interpretations (Horn, 1989; Rooth, 1992). 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).

Scope of negation detection has received a lot of attention (Özgür and Radev, 2009; Packard et al., 2014), mostly using two corpora: BioScope (Szarvas et al., 2008), and CD-SCO (Morante and Daelemans, 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).

Identifying the focus of negation is generally more challenging than the scope. The challenge lies on determining which tokens within the scope are *intended* to be negated. The largest corpus to date is PB-FOC, which was released as part of the *SEM-2012 Shared Task (Morante and Blanco, 2012). PB-FOC annotates the semantic role most likely to be the focus in the 3,993 negation in PropBank (Palmer et al., 2005). Anand and Martell (2012) refine PB-FOC and argue that 27.4% of negations with a focus annotated in PB-FOC do not actually have a focus. Sarabi and Blanco (2016) present a complementary approach grounded on syntactic dependencies. All of these efforts identify the tokens that are the focus of negation. We build upon them and generate actual positive interpretations from negation.

4 Corpus Creation

This section details our data collection and annotation effort. We follow 5 steps. First, we describe the source corpus. Second, we outline the procedure to select negations so that the annota-

		#	%
# negations	1 negation	69,365	93.3%
	≥2 negations	4,981	7.7%
# tokens	≤5	683	0.9%
	>5 & ≤25	50,070	67.3%
	>25	23,593	31.7%

Table 2: Basic counts for sentences containing negation in Simple Wikipedia.

			#	%
Negation Types	Verbal	root	24,125	32.4%
		not_root	31,386	42.2%
	Nominal		11,003	14.8%
	Adjectival		2,325	3.1%
	Other		5,507	7.4%
	All		74,346	100.0%

Table 3: Distribution of negation type in Simple Wikipedia.

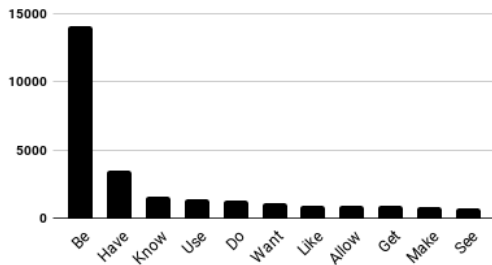


Figure 1: The most frequent negated verb lemmas in Simple Wikipedia.

tion effort is feasible. Third, we discuss the steps to automatically generate potential positive interpretations. Fourth, we detail the annotation effort to rewrite placeholders in the potential positive interpretations to generate actual positive interpretations. Fifth, we present the final validation strategy to ensure quality of the final corpus.

4.1 Selecting the Source Corpus: Simple Wikipedia

We chose to work with Simple Wikipedia texts.³ Simple Wikipedia is a version of Wikipedia that is written in basic English. Compared to regular Wikipedia, articles in Simple Wikipedia use simpler words, shorter sentences, and simple grammar. These characteristics help us to reduce the overhead of dealing with complex sentences and leads to a more realistic learning task. We process Simple Wikipedia with spaCy (Honnibal and Johnson, 2015) to obtain part-of-speech tags and dependency trees. Inspired by Fancellu et al.

³Version 2018-03-01; available at <https://dumps.wikimedia.org/simplewiki/>

(2016), we identify sentences containing negation using the following cues: *n't*, *not*, *never*, *no*, *nothing*, *nobody*, *none*, *nowhere*. Note that this method selects negations that would be discarded if we relied only on dependency type *neg*.

Table 2 shows basic counts for sentences containing at least one negation in Simple Wikipedia. 93% of them contain only one negation, and 67% have medium length (between 6 to 25 tokens). Table 3 categorizes the Simple Wikipedia negations based on their type. We identify negation types using the part-of-speech tag of the syntactic head of the negation cue, i.e., the syntactic parent or governor of the negation cue. More than 70% of the negations in Simple Wikipedia are verbal negations, and the verb is the root of the dependency tree in 44% of them. Finally, Figure 1 shows the most frequent verbal negations in Simple Wikipedia. We observe that many verbs and in particular the verb *to be* are very frequent, and there is a long tail of (relatively) infrequent verbs.

4.2 Selecting Negations

Working with all negation types in Simple Wikipedia is out of the scope of this paper. After doing pilot annotations and manual examination, we decided to limit the negation types grounded on the counts presented in Section 4.1. Table 4 summarizes the filters and the number of negations that remain after running each filter. We apply sequentially five filters (Filters 1–5) on negations and four filters (Filters 6–9) on sentences. Filter 1 discards non-verbal negations (recall that 74.6% of negations are verbal, Table 3). Filter 2 discards those verbal negations which are not the root of the dependency tree. Filter 3 discards infrequent verbal negations, more specifically, those whose verbs occurred less than five times. Filter 4 caps the number of verbal negations per verb to 200 negations to increase verb coverage (recall that some verbs are negated very frequently, Figure 1). Filter 5 discards verbal negations with part-of-speech tag *interjection* (less than 1%, e.g., *They said “no” to his offer*). Filter 6 discards sentences whose length is not greater than five tokens and less than 26 tokens (recall that most sentences containing negation satisfy this filter: 67.3%, Table 2). Filter 7 discards sentences with more than one verbal negation (93% of sentences containing negation only contain one, Table 2). Filter 8 discards negated sentences in question form (i.e., the first

	Filters	#	%
Initial #	-	74,346	100.0
Negations	F1: non_verbal	58,756	79.0
	F2: not_root	27,370	36.8
	F3: infrequent	18,618	25.0
	F4: limit_freq	16,185	21.7
	F5: intj	16,136	21.1
Sentences	F6: limit_len	13,859	19.1
	F7: multi_negs	10,614	14.8
	F8: questions	10,476	14.6
	F9: other	7,469	10.1

Table 4: Initial number of negations in Simple Wikipedia, and how many remain after each filter.

token has any of the following part-of-speech tags: WDT, WP, WRB). Filter 9 discards sentences that include any of the following tokens: *because, until, but, if, except*. The final dataset consists of 7,469 negations, which are approximately 10% of negations in Simple Wikipedia.

4.3 Generating Potential Positive Interpretations

We convert each negation into its positive counterpart in four steps following the rules by Huddleston and Pullum (2002): remove the negation cue, remove auxiliaries, fix third-person singular and past tense, and rewrite negatively-oriented polarity-sensitive items. These steps can be implemented using straightforward regular expressions. For example, the positive counterpart of *The seller did not ship the right part*, is *The seller shipped the right parts*. Then, we automatically generate all plausible positive interpretations of the negation by traversing the dependency tree and selecting all direct dependents of the negated verb. We filter out subtrees whose syntactic dependency is *aux*, *auxpass*, *punct* (auxiliary, passive auxiliary and punctuation). We also exclude the verb. These exceptions were defined after manual examination of several examples. Finally, we replace the selected subtrees with a placeholder.

Table 5 shows the number of negations depending on how many positive interpretations are generated. We generate two or more potential positive interpretations for over 84% of negations.

4.4 Rewriting Placeholders

In order to rewrite placeholders in potential positive interpretations and collect actual positive interpretations, we implement an annotation inter-

# pot. positive interpretations	#	%
1	1,132	15.53
2	3,723	51.07
3	2,004	27.49
4	437	5.99
5	29	0.39
6	3	0.04
All	7,469	100.00

Table 5: Distribution of negations by number of potential positive interpretations generated.

face using Amazon Mechanical Turk Sandbox.⁴ This rewriting process was done in-house by one linguistics student. A second annotator validated the rewrites independently (Section 4.5).

Each negation along with its context and all its potential positive interpretations are grouped into a Human Intelligence Task (HIT) for annotation purposes. Each HIT presents a set of instructions to the annotator along with examples. Potential positive interpretations are presented in consecutive rows, and each token in a cell. The placeholders generated in Section 4.3 are presented as blank cells and the annotator fills the blanks (or, in other words, the annotator rewrites placeholders) based on the context around the negation or world knowledge. A sample HIT along with the answers collected is shown in Figure 2.

In the rest of the paper, we use *unknown* answer to refer to placeholders for which the annotator cannot find a rewriting. We divide *unknown* answers into *invalid* and *not specified*, and ask the annotator to distinguish between them. *Invalid* is used to refer to placeholders that cannot be rewritten. *Not specified* describes placeholders that hypothetically can be rewritten but the answer is unknown given the context. We also provide an extra empty box at the bottom of the interface for additional positive interpretations. If the annotator cannot find any answers for the rewrites, she can write a positive interpretation from scratch.

4.5 Validating Positive Interpretations

In order to validate the rewrites of placeholders and resulting positive interpretations (Section 4.4), a second annotator validates them. We create a similar interface to the one in Figure 2, but this time we only show the negation in context (*Text* in Figure 2), and one positive interpretation at a time (i.e., potential positive interpretation for which the

⁴<https://requester.mturk.com/developer/sandbox>

Text:
 Breakup**The word "breakup" is an idiom. It is when an intimate relationship ends. The term is not used for relationships that end because of death. **Marriages which end are not normally called breakups.**
<https://simple.wikipedia.org/wiki?curid=455908>

Reminder

			are	normally	called	breakups	.
Marriages	which	end	are		called	breakups	.
Marriages	which	end	are	normally	called		.

Additional Positive Interpretation:

Figure 2: Sample negation along with its context and automatically-generated potential positive interpretations. The annotation process reveals three positive interpretations: “*Relationships* that end are normally called breakups,” “Marriages which end are *rarely* called breakups,” and “Marriages which end are normally called *divorce*.”

placeholder was rewritten). The annotator determines correctness and novelty as follows.

Correctness measures whether a positive interpretation is true given the negation in context. It is measured using the following scale:

1. After reading the *text*, it is clear that the positive interpretation is false.
2. After reading the *text*, the positive interpretation is probably false, but I am not sure.
3. After reading the *text*, the positive interpretation is probably true, but I am not sure.
4. After reading the *text*, it is clear that the positive interpretation is true.

Novelty measures whether the meaning conveyed by a positive interpretation is already explicitly stated in the *text*, and it is measured using the following numeric scale:

1. The positive interpretation is stated explicitly in the *text* with the very same words. I could copy and paste chunks from *text* and get the positive interpretation.
2. The positive interpretation is not stated in the *text* with the same words. The positive interpretation and the *text* have synonyms in common, but I could not get the positive interpretation simply copying and pasting from *text*.
3. The positive interpretation is not stated in the *text* with the same words. Additionally, there are few synonyms in common between the positive interpretation and *text*.

5 Corpus Analysis

The procedure described in Section 4 generates 15,875 potential positive interpretations from 7,469 negations. Out of all potential positive interpretations, we rewrite 3,831 with an actual answer and annotate 12,044 with an unknown answer

	Known		Unknown				Total
	#	%	not spec.		invalid		
			#	%	#	%	
prep	953	29	2089	63	262	8	3305
dobj	763	30	1652	65	121	5	2537
advmod	700	50	629	45	68	5	1398
nsubj	511	11	3860	83	295	6	4667
nsubjpass	212	14	1267	82	62	4	1542
ccomp	207	34	357	58	52	8	617
xcomp	168	38	248	57	21	5	438
advcl	131	18	534	73	63	9	729
agent	60	29	142	68	5	2	208
Other	128	27	250	53	99	20	477

Table 6: Corpus analysis. For each dependency type (left column), we show the number of potential positive interpretations generated with known and unknown rewrites (and *not specified* and *invalid* rewrites).

(11,030 *not specified* and 1,014 *invalid*). We also rewrite a new positive interpretation from scratch for 2,158 negations for which we cannot find any actual rewrites. Overall, we rewrite 5,989 positive interpretations for 5,770 unique negations. In other words, the procedure in Section 4 yields a positive interpretation for 77% of negations.

Table 6 shows the distribution of *known* vs *unknown* rewrites per dependency type, where *dependency type* refers to the dependency type from the selected subtree of the verb to the verb itself. Out of all dependency types, *advmod* and *xcomp* (adverbial modifier and open clausal complement respectively) have the highest ratios of known rewrites, and *nsubj* (nominal subject) has the most unknown answers. In other words, the easiest placeholders to rewrite are those whose syntactic function is adverbial modifier or open clausal complement, and the most challenging are those whose syntactic function is nominal subject.

To understand high-level characteristics of negations and their positive interpretations beyond

Category	Subcat.	Examples	%
Quantities	specific	Members of Congress cannot serve for more than <i>three out of any six years</i> .	3%
	abstract	<i>Many</i> do not use their real names, as Everett does.	22%
Times	actual	<i>Since 2012</i> , this channel never goes off the air during the day.	4%
	abstract	Rabbits should not be bred <i>too early</i> though.	11%
Objects	-	It does not need <i>sunlight</i> to grow and can stay in the same pot for many years.	9%
Adjectives	-	Crops did not grow <i>as well</i> when they were close together.	27%
Proper nouns	-	<i>Cosworth</i> does not currently provide engines to any American open wheel racing series.	2%
Others	-	<i>The mass number</i> is not shown on the periodic table.	22%

Table 7: Categories and subcategories discovered in a sample of 100 negations and all their positive interpretations.

dependency types, we explore a random sample of 100 negations and all their positive interpretations. We discover six major categories (quantities, times, objects, adjectives, proper nouns and others) and 4 subcategories (Table 7):

- The first category is *quantities* and includes both *specific* and *abstract* quantities. An example of abstract quantity is *Many do not use their real names, as Everett does* and its corresponding positive interpretation *Few use their real names, as Everett does*. A fourth of positive interpretations in the sample were obtained after rewriting quantities.
- The second category is *time* and includes both *actual* and *abstract* times. An example of actual time is *Since 2012, this channel never goes off the air during the day* and its positive interpretation *Before 2012, this channel went off the air during the day*. 15% of positive interpretations in the sample were obtained rewriting temporal expressions.
- The third category is *objects* and refers to positive interpretations obtained by rewriting verbal objects. An example is *It does not need sunlight to grow* and its positive interpretation *It needs water to grow*. 9% of positive interpretations in the sample were obtained after rewriting the verbal objects.
- The fourth category is *adjectives* and refers to positive interpretations obtained by rewriting adjectives. An example is *Crops did not grow as well when they were close together* and its positive interpretation *Crops grew poorly when they were close together*. 27% of positive interpretations in the sample were obtained after rewriting adjectives.
- The fifth category is *proper nouns*. An example is *Cosworth does not currently provide engines to any American open wheel racing series* and its positive interpretation *IndyCar Series currently provide engines to*

American open wheel racing series. 2% of positive interpretation in the sample were obtained after rewriting proper nouns.

5.1 Annotation Quality

To assess the quality of the rewrites and positive interpretations, we ask a second annotator to validate them based on two criteria: correctness and novelty (Section 4.5). Recall that correctness ranges from 1 (minimum) to 4 (maximum) and novelty from 1 (minimum) to 3 (maximum). We assess novelty only if positive interpretations are correct (correctness scores 3 or 4). Figure 3 reports the validation results. Out of all positive interpretations obtained during the annotation process, 90% are either correct (77%) or probably correct (13%) (correctness scores 4 and 3), and 95% of them are either very novel (52%) or novel (43%). This validation scores mean not only that positive interpretations are sound given the original negation (correctness score), but also that they are not explicitly stated in the context and thus reveal implicit meaning (novelty score).

5.2 Annotation Examples

Table 8 presents three negations, all potential positive interpretations, and manual annotations along with the correctness and novelty scores.

Example (1) is a simple negated clause. The procedure described in Section 4.3 generates four potential positive interpretations, and three of them were rewritten. Given *Phosgene usually does not cause its worst effects right away* and its context, the following positive interpretations are deemed correct (correctness = 4) with different degrees of novelty (2, 3 and 1 respectively): *Phosgene rarely causes its worst effects right away* (Interpretation 1.2), *Phosgene usually causes mild effects right away* (Interpretation 1.3), and *Phosgene usually causes its worst effects 12 hours after a person breathes it in* (Interpretation 1.4). Note that

1	Context: Phosgene can be a liquid or a gas. As a gas, it is heavier than air, so it can stay near the ground (where people can breathe it in for long periods of time). It smells like freshly cut grass or moldy hay. Along with being a choking agent, phosgene is also a blood agent. This means it keeps oxygen from getting into the body’s cells. Without oxygen, a person’s cells will die, and the person will suffocate. <i>Phosgene usually does not cause its worst effects right away.</i> The worst symptoms do not happen until 12 hours after a person breathed in phosgene. The person usually dies within 24 to 48 hours.		
	Sentence containing negation: Phosgene usually does not cause its worst effects right away.		
		Correctness	Novelty
	- Interpretation 1.1: [NS] usually causes its worst effects right away.	-	-
	- Interpretation 1.2: Phosgene <i>rarely</i> causes its worst effects right away.	4	2
- Interpretation 1.3: Phosgene usually causes <i>mild effects</i> right away.	4	3	
- Interpretation 1.4: Phosgene usually causes its worst effects <i>12 hours after a person breathes it in.</i>	4	1	
2	Context: Hungary uses Central European Time (CET) which is 1 hour ahead of Coordinated Universal Time (UTC+1). <i>Hungary has not observed summer time since 1916.</i>		
	Sentence containing negation: Hungary has not observed summer time since 1916.		
		Correctness	Novelty
	- Interpretation 2.1: [NS] has observed summer time since 1916.	-	-
	- Interpretation 2.2: Hungary has observed <i>Central European Time</i> since 1916.	4	1
- Interpretation 2.3: Hungary has observed summer time <i>prior to 1916.</i>	4	2	
3	Sentence containing negation: This does not stop him from finding ways to try to get more money.		
		Correctness	Novelty
	- Interpretation 3.1: This stops [NS] from finding ways to try to get more money.	-	-
	- Interpretation 3.2: This stops him [NS].	-	-
	- Addtl. Interpretation: He’s always trying to get more money despite being rich.	4	3

Table 8: Sentences containing negation (and context if relevant to obtain positive interpretations), all automatically-generated positive interpretations, positive interpretations manually annotated (italics indicate placeholder rewrites), and validation scores (correctness and novelty).

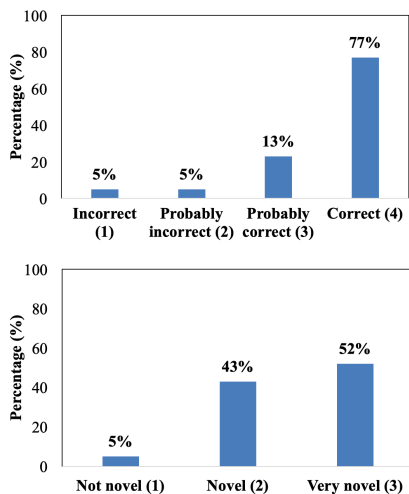


Figure 3: Distribution of correctness (top) and novelty (bottom) scores in our corpus.

Interpretation 1.1 is most likely correct, but context does not provide clues about which chemicals cause their worst effects right away and thus it is annotated *not specified* (NS).

Example (2) has three potential positive interpretations, and we rewrite two of them. Note that Interpretation 2.2, *Hungary has observed Central European Time since 1916*, is correct but not novel because it is explicitly stated in the context. Interpretation 2.3 is correct but received novelty score

of 2 because it only replaces *since* with *prior to*.

Example (3) shows an example in which rewriting placeholders is not successful. The additional interpretation, however, reveals that *He* has the intention of getting more money. Context, which is not shown in Table 8, support the correctness and validation scores (e.g., *He* is wealthy).

6 Experiments

The task of generating positive interpretations from a sentence containing negation can be approached with sequence-to-sequence (seq2seq) models (input: sentence containing negation, output: positive interpretation). In this section, we present baseline results with existing seq2seq models. Specifically, we experiment with a basic seq2seq model (Cho et al., 2014), two seq2seq models with attention (Luong et al., 2015; Bahdanau et al., 2014), and Google’s neural machine translation (NMT) system (Wu et al., 2016), which is also seq2seq model with attention and arguably the most complex. We acknowledge that these systems are usually trained with orders of magnitude more examples, and comparing them when trained with our fairly small corpus may be unfair because they were designed for other tasks. Our goal is not to obtain the best results possible, but rather provide baseline results for our task and corpus.

Models	Short Sentences			Long Sentences		
	BLEU	Gram.	Corr.	BLEU	Gram.	Corr.
seq2seq (basic) (Cho et al., 2014)	10.31	23%	5%	2.13	6%	1%
seq2seq + attention (Bahdanau et al., 2014)	20.51	65%	22%	9.20	41%	12%
seq2seq + attention (Luong et al., 2015)	28.08	68%	30%	14.53	51%	19%
seq2seq + Google’s NMT attention (Wu et al., 2016)	12.54	42%	15%	4.40	12%	3%

Table 9: Results (BLEU-4, grammaticality and correctness) obtained with the test set.

The 3,831 negations become source sentences and the correct positive interpretations become target sentences. We randomly select 100 short sentences (up to 12 tokens) and 100 long sentences (over 12 tokens) for testing, 200 sentences for development, and the remainder for training. All positive interpretations collected from a negation are assigned to the testing, development or training splits in order to ensure a more realistic scenario.

Evaluation and Results. We use three metrics to evaluate the models: BLEU-4, correctness and grammaticality. BLEU-4 is automated, convenient and useful for development purposes. While larger BLEU-4 scores generally indicate better correctness and grammaticality scores, we do not observe a linear correlation (Table 9). Correctness is measured manually with the scale presented in Section 4.5. Finally, grammaticality is measured manually using the following numeric scale:

1. The sentence is not grammatical at all, e.g., it does not contain a verb.
2. The sentence is mostly ungrammatical, e.g., it contains a verb but the word order is wrong.
3. The sentence has a few grammatical issues, e.g., the subject-verb agreement is wrong, missing punctuation.
4. The sentence is grammatically correct (regardless of its correctness).

Table 9 shows the results. In general terms, results are better for short sentences than long ones. This is not surprising given the small size of our corpus. The basic seq2seq model performs poorly: it barely generates any correct positive interpretations, and most are ungrammatical. Adding attention performs better. The best results are with the system by Luong et al. (2017): 30% of the short positive interpretations generated are correct, and 68% grammatical. We believe Google’s NMT performs the worst because of the small corpus.

We also conduct a manual analysis of the correct positive interpretations generated by the best system. Following with the categories described in Section 5 and Table 7, 37% of them belong to the *adjectives* category, 27% to *abstract quanti-*

ties, 17% to *objects*, and 10% to *abstract time*.

7 Conclusions

We have presented a corpus of negations and their positive interpretations. Positive interpretations do not contain negations, range from implicatures to entailments, and are intuitively understood by nonexperts when reading the negations. We work with verbal negations selected from Simple Wikipedia, automatically generate potential positive interpretation by replacing subtrees with placeholders, and manually collect rewrites for the placeholders in order to obtain actual positive interpretations. This strategy yields positive interpretations for 77% of negations, and manual validation step ensures both correctness and novelty.

Neural machine translation struggles with negation, and natural language inference benchmarks do not account for the intricacies of negation (Section 2). While small, we believe the corpus presented here is a step towards enabling natural language understanding when negation is present.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants Nos. 1734730, 1832267 and 1845757. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The Titan Xp used for this research was donated by the NVIDIA Corporation.

References

- 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*. Association for Computational Linguistics, Atlanta, Georgia, USA, pages 32–43. <http://www.aclweb.org/anthology/S13-1004>.

- 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*. Association for Computational Linguistics, Stroudsburg, PA, USA, ExProM '12, pages 65–69. <http://dl.acm.org/citation.cfm?id=2392701.2392709>.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv e-prints* abs/1409.0473. <https://arxiv.org/abs/1409.0473>.
- Luisa Bentivogli, Arianna Bisazza, Mauro Cettolo, and Marcello Federico. 2016. Neural versus phrase-based machine translation quality: a case study. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 257–267. <https://doi.org/10.18653/v1/D16-1025>.
- Simon Blackburn. 2008. *The Oxford Dictionary of Philosophy*. Oxford University Press. <http://www.oxfordreference.com/10.1093/acref/9780199541430.001.0001/acref-9780199541430>.
- Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 632–642. <https://doi.org/10.18653/v1/D15-1075>.
- Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2017. Recurrent neural network-based sentence encoder with gated attention for natural language inference. In *Proceedings of the 2nd Workshop on Evaluating Vector Space Representations for NLP*. Association for Computational Linguistics, Copenhagen, Denmark, pages 36–40. <http://www.aclweb.org/anthology/W17-5307>.
- Kyunghyun Cho, Bart van Merriënboer, Çağlar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, pages 1724–1734. <http://www.aclweb.org/anthology/D14-1179>.
- Federico Fancellu, Adam Lopez, and Bonnie Webber. 2016. Neural networks for negation scope detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, pages 495–504. <http://www.aclweb.org/anthology/P16-1047>.
- Matthew Honnibal and Mark Johnson. 2015. An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pages 1373–1378. <https://aclweb.org/anthology/D/D15/D15-1162>.
- Laurence R. Horn. 1989. *A natural history of negation*. Chicago University Press, Chicago.
- Laurence R. Horn and Heinrich Wansing. 2015. Negation. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2015 edition.
- Rodney D. Huddleston and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press.
- 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)*. Coling 2010 Organizing Committee, Beijing, China, pages 671–679. <http://www.aclweb.org/anthology-new/C/C10/C10-1076.bib>.
- Minh-Thang Luong, Eugene Brevdo, and Rui Zhao. 2017. Neural machine translation (seq2seq) tutorial. <https://github.com/tensorflow/nmt>.
- Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, pages 1412–1421. <https://doi.org/10.18653/v1/D15-1166>.
- Ruslan Mitkov. 2005. *The Oxford handbook of computational linguistics*. Oxford University Press.
- R. Morante and W. Daelemans. 2012. Annotating modality and negation for a machine reading evaluation. In *CLEF 2012 Evaluation Labs and Workshop Online Working Notes*.
- 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 - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics, Stroudsburg, PA, USA, SemEval '12, pages 265–274. <http://dl.acm.org/citation.cfm?id=2387636.2387679>.

- 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*. Association for Computational Linguistics, Singapore, pages 1398–1407. <http://www.aclweb.org/anthology-new/D/D09/D09-1145.bib>.
- Woodley Packard, Emily M. Bender, Jonathon Read, Stephan Oepen, and Rebecca Dridan. 2014. Simple negation scope resolution through deep parsing: A semantic solution to a semantic problem. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Baltimore, Maryland, pages 69–78. <http://www.aclweb.org/anthology/P14-1007>.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics* 31(1):71–106.
- Randolph Quirk, Sidney Greenbaum, and Geoffrey Leech. 2000. *A comprehensive grammar of the English language*. Longman, London.
- Mats Rooth. 1985. *Association with focus*. Ph.D. thesis.
- Mats Rooth. 1992. A theory of focus interpretation. *Natural language semantics* 1(1):75–116.
- Zahra Sarabi and Eduardo Blanco. 2016. Understanding negation in positive terms using syntactic dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Austin, Texas, pages 1108–1118. <https://aclweb.org/anthology/D16-1119>.
- György Szarvas, Veronika Vincze, Richárd Farkas, and János Csirik. 2008. The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. In *Proceedings of BioNLP 2008*. ACL, Columbus, Ohio, USA, pages 38–45.
- Ton van der Wouden. 1997. *Negative contexts: collocation, polarity, and multiple negation*. Routledge, London.
- Erik Velldal, Lilja Ovreliid, Jonathon Read, and Stephan Oepen. 2012. Speculation and negation: Rules, rankers, and the role of syntax. *Comput. Linguist.* 38(2):369–410. https://doi.org/10.1162/COLI_a_00126.
- Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Gregory S.
- Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google’s neural machine translation system: Bridging the gap between human and machine translation. *CoRR* abs/1609.08144.