SLSA: A Sentiment Lexicon for Standard Arabic

Ramy Eskander and Owen Rambow
Center for Computational Learning Systems
Columbia University
{rnd2110,ocr2101}@columbia.edu

Abstract

Sentiment analysis has been a major area of interest, for which the existence of high-quality resources is crucial. In Arabic, there is a reasonable number of sentiment lexicons but with major deficiencies. The paper presents a large-scale Standard Arabic Sentiment Lexicon (SLSA) that is publicly available for free and avoids the deficiencies in the current resources. SLSA has the highest up-to-date reported coverage. The construction of SLSA is based on linking the lexicon of AraMorph with SentiWordNet along with a few heuristics and powerful back-off. SLSA shows a relative improvement of 37.8% over a state-of-the-art lexicon when tested for accuracy. It also outperforms it by an absolute 3.5% of F1-score when tested for sentiment analysis.

1 Introduction

Sentiment analysis is the process of identifying and extracting subjective information using Natural Language Processing (NLP). It helps identifying opinions and extracting relevant information that lies behind the analyzed data. Sentiment analysis has received enormous interest in NLP, and in particular in the context of web content. This includes social media, blogs, discussions, reviews and advertisement.

While there has been extensive work on sentiment analysis in English and other languages of interest, less work has been done for Arabic. A major concern in Arabic NLP is the morphological complexity of the language along with the limited number of resources, corpora in particular.

The goal of this work is to build a publicly available large-scale Sentiment Lexicon for Standard Arabic (SLSA). For every lemma and part-of-speech (POS) combination that exists in a large Standard Arabic lexicon, SLSA assigns the scores of three sentiment labels: positive, negative and objective, in addition to the English gloss. The positive and negative scores range between zero and one, while the objective score is defined as *1 - (positive score + negative score)*.

The existence of SLSA is valuable to the field of Arabic sentiment analysis, which is expected to receive considerable focus during the current decade. SLSA is the first sentiment lexicon for Arabic to combine the following four strengths.

High coverage SLSA lists the sentiment of about 35,000 lemma and POS combinations, which is the highest coverage reported for Standard Arabic sentiment lexicons.

High quality Unlike many of the current lexicons whose construction is based on semi-supervised learning and heuristic-based approaches, SLSA is not constructed via machine learning models, while the use of heuristics is minimal.

Richness As opposed to sparse surface-based lexicons, SLSA is a lemma-based resource that attaches POS and English gloss information to each lemma, where the information of a lemma is applicable to its inflected forms. This makes the lexicon more useful when used by other research.

Public Availability SLSA is based on free resources and is publicly available for free.[1]

2 Related Work

Work on building Arabic sentiment lexicons mainly falls into two categories: 1) linking an Arabic sentiment lexicon with an English one, and 2) applying semi-supervised or supervised learning techniques on Arabic resources. We summarize these two types in turn.

We start with a survey of work based on translation, which our work falls into as well. The most similar work to the one presented in this paper is ArSenL (Badaro et al., 2014). ArSenL is considered the first publicly available large-scale Standard Arabic sentiment lexicon. It was constructed using a combination of SentiWordNet (Baccianella et al., 2010), Arabic WordNet (Black et al., 2006) and SAMA (Graff et al., 2009). Ar-

---

[1] The lexicon is available at http://volta.ldeo.columbia.edu/~rambow/slsa.html

SenL outperforms the state-of-the-art Arabic sentiment lexicons. However, we show that SLSA has better coverage and quality. Moreover, ArSenL uses SAMA which is not publicly available for free, as opposed to SLSA which is based on free resources.

Another similar work to ArSenL is the resource developed by Alhazmi et al. (2013). They linked the Arabic WordNet to SentiWordNet via the provided synset offset information. However, the constructed lexicon has a limited coverage of nearly 10K lemmas, which makes it not very useful for further applications.

Abdul-Mageed and Diab (2014) presented SANA, a subjectivity and sentiment lexicon for Arabic. The lexicon combines pre-existing lexicons and involves automatic machine translation, manual annotations and gloss matching across several resources such as THARWA (Diab et al., 2014) and SAMA. SANA includes about 225K entries, where many of them are duplicates, inflected or not diacritized, which makes the resource noisy and less useable. Additionally, the automatic translation does not utilize the POS information, which affects the quality of the resource.

Other work that follows the translation approach includes the one presented by El-Halees (2011) where SentiStrength (Thelwall et al., 2010) was translated using a dictionary along with manual correction. Another instance is SIFAAT (Abdul-Mageed and Diab, 2012), an earlier version of SANA but with more reliance on translation. Another lexicon was built by Elarnaoty et al. (2012) who manually translated the MPQA lexicon (Wilson et al., 2005). The common aspect among those resources is the lack of adequate coverage and quality.

Mobarz et al. (2011) created a sentiment Arabic lexical Semantic Database (SentiRDI) by using a dictionary-based approach. The database has many inflected forms, i.e., it is not lemma-based. Moreover, the authors reported insufficient quality and plan to try other alternatives.

We now turn to work based entirely on Arabic resources. Mahyoub et al. (2014) created an Arabic sentiment lexicon that assigns sentiment scores to the words in Arabic WordNet using a lexicon-based approach. The lexicon was initially based on a few words and then expanded by exploiting synset relations in a semi-supervised learning manner. However, the lexicon is limited to about 23k lemmas and is not publicly available.

Another Arabic sentiment lexicon was created by Elhawary and Elfeky (2010). The lexicon was built using a similarity graph where the edges have similarity scores. A major drawback is the low coverage of the lexicon. Moreover, expanding the graph requires a huge corpus with polarity and semantic annotations and adds more sparsity.

3 Approach

Following the example of ArSenL (Badaro et al., 2014), SLSA is constructed by linking the lexicon of an Arabic morphological analyzer with SentiWordNet (Baccianella et al., 2010). Unlike ArSenL, SLSA uses AraMorph (Buckwalter, 2004), a morphological analyzer for Standard Arabic. An AraMorph entry represents a morpheme and contains the surface, lemma, part of speech (POS), and gloss information. The gloss information consists of a list of gloss terms, each of which contains one or more words (such as "*time limit / end*"). On the other side, SentiWordNet is a large-scale sentiment lexicon for English that assigns sentiment scores (positive, negative and objective) to the synsets in English WordNet (Miller et al., 1990) along with the POS and gloss information. Upon linking the two resources, the sentiment scores in SentiWordNet are applied to the entries of AraMorph to construct SLSA. The question this paper addresses is how to link these two resources, and we present a new linking algorithm compared to that used by ArSenL, with improved performance.

3.1 Preparing the Resources

It might seem intuitive to join the entries of AraMorph and SentiWordNet based on their glosses, but this does not work as expected. AraMorph and SentiWordNet were developed for different reasons and have different gloss structures (synonyms in AraMorph versus detailed descriptions in SentiWordNet). Mapping the glosses is one of the major bottlenecks in ArSenL, which is not able to find a match for 24% of the entries in SAMA. Instead, we link the two resources by relating the glosses of AraMorph to the synset terms in SentiWordNet. Additionally, we take POS into consideration as the glosses and synset terms might not be enough to disambiguate an entry. Next, we discuss the preparation steps that allow for the linking of the resources.

Cleaning-up AraMorph Some POS and lemma decisions in AraMorph are erroneous or not optimal. For example, some entries are assigned wrong POS tags, such as the NO_FUNC cases, or have inconsistent spellings of the lemmas. Also, some adverbs are redundant as they have the same lemma as an adjective. Accordingly, we cleaned up AraMorph in a way that allows for a better linking with SentiWordNet. The

cleaned-up AraMorph is closer to SAMA (used in ArSenL), which is itself a modified version of AraMorph. Practically, SAMA can replace AraMorph. However, the integration of AraMorph allows the lexicon to be publicly available for free, which SAMA prohibits.

Gloss Normalization Since the entries in AraMorph are bound to stems, the English glosses are inflected for number. As a result, we lemmatize the English glosses in AraMorph in order to be able to match to the synset terms in SentiWordNet. The lemmatization is done using Stanford CoreNLP Natural Language Processing Toolkit (Manning et al., 2014). Additionally, we remove from the glosses any descriptive text between parentheses, as well as the stop words *be*, *a*, *an* and *the* (unless *be* is the actual lemma of the AraMorph gloss). Moreover, if any of the lemmatized words in an AraMorph gloss does not match any of the synset terms in SentiWordNet and has a regular morphological derivation, the effect of the derivation is removed if the removal results in an existing synset term, e.g., *voluntariness* is converted to *voluntary* and *orientalization* becomes *orientalize*. We created the list of the derivational patterns manually by examining AraMorph glosses.

POS Mapping AraMorph has a rich POS tagset, while SentiWordNet has only four tags corresponding to nouns, adjectives, adverbs and verbs. Accordingly, AraMorph POS tags are mapped to the four tags in SentiWordNet. Some AraMorph POS types, such as particles, pronouns and prepositions, do not map to any SentiWordNet tags, and we exclude them as they have zero polarity scores, by definition.

AraMorph Rearrangement We collapse all the entries in AraMorph that have the same (lemma, POS) pair, and the English glosses become the union of the normalized glosses before the collapse. For example, a lemma might appear in two entries with two POS tags; *VERB_PERFECT* and *VERB_IMPERFECT*. After the preparation, the POS tags in both entries become the same (*VERB*), and the two entries collapse into one entry whose gloss is the union of the lemmatized past-tense and present-tense glosses. Figure 1 shows a sample of AraMorph before and after the preparation process (the Arabic transliteration is in the Buckwalter scheme (Buckwalter, 2004)).

SentiWordNet Rearrangement We extract all the unique combinations of synset terms and POS tags in SentiWordNet, while the indices of the synset terms are stripped off. However, since a synset term might appear in different synsets un-

Stem	English Gloss	POS	Lemma
baliy~	tribulation/affliction	NOUN	baliy~ap_1
balAyA	tribulations/afflictions	NOUN	baliy~ap_1
balA	afflict/test	VERB_PERFECT	balA-u_1
balaw	afflict/test	VERB_PERFECT	balA-u_1
bal	afflict/test	VERB_PERFECT	balA-u_1
boluw	afflict/test	VERB_IMPERFECT	balA-u_1
bol	afflict/test	VERB_IMPERFECT	balA-u_1
bolaY	be afflicted/be tested	VERB_IMPERFECT	balA-u_1

Original AraMorph

Lemma	POS	English Gloss
baliy~ap_1	NOUN	affliction/tribulation
balA_1	VERB	afflict/test

Processed AraMorph

Figure 1: A sample of the original AraMorph (the upper table) and the processed version (the lower table). The glosses are normalized, while the POS tags are mapped to the tags in SentiWordNet. The entries of the same POS and lemma combinations are then collapsed, where their gloss becomes the union of the normalized glosses in the collapsed entries.

der the same POS with different indices and sentiment scores, the sentiment scores of an extracted entry is calculated as the average of all the sentiment scores that appear with the corresponding synset term and POS. Figure 2 shows a sample of SentiWordNet before and after the preparation process.

POS	ID	+ve	-ve	Synset Terms	Gloss
n	07305234	0	0.625	affliction#3	a cause of great suffering and distress
n	14213199	0	0.625	affliction#2	a condition of suffering or distress...
n	14477342	0	0.625	affliction#1	a state of great suffering and distress...
v	00259927	0	0.875	smite#3 afflict#2	cause physical pain or suffering in
v	01797730	0.125	0.625	afflict#1	cause great unhappiness for

Original SentiWordNet

Synset Term	POS	+ve	-ve
affliction	NOUN	0	0.625
afflict	VERB	0.0625	0.75

Processed SentiWordNet

Figure 2: A sample of the original SentiWordNet (the upper table) and the processed version (the lower table). The unique combinations of the synset terms and POS tags are extracted, where the sentiment scores of the extracted entries are the average of the scores in the contributing ones.

3.2 Linking the Resources

We start out by creating a link between an AraMorph entry and a SentiWordNet entry if any of the AraMorph one-word gloss terms and the POS match the SentiWordNet. Upon linking, we assign the AraMorph entry the sentiment scores of the matching one in SentiWordNet. The linking condition above applies successfully to 83.6% of the entries in AraMorph. If the condition does not apply, we relax it to allow for a more lenient POS agreement where *NOUN* and *ADJ* POS tags are used interchangeably, while the *VERB* entries in AraMorph become matchable with the *ADJ* ones in SentiWordNet. The reasons behind the decisions above are that AraMorph has hundreds of cases where the same lemma appears as *NOUN* and *ADJ*, while it is frequent that AraMorph assigns an adjectival gloss (preceded

by *be*) to *VERB* entries. The relaxed condition enables linking an additional 6.7% of AraMorph entries. If the relaxed condition is still not applicable for an AraMorph entry, the linking condition becomes more lenient by completely ignoring the POS agreement. The sentiment scores in that case become the average of the sentiment scores of the corresponding synset term across all the POS types. The latter condition allows matching additional 0.6% of AraMorph entries.

It might happen that none of the one-word gloss terms matches a synset term, or the gloss does not have any one-word gloss terms. In such a case, we consider multi-word gloss terms. We first remove the stop words, and then we test the relaxed condition on each word separately, starting with the shortest terms first. The process succeeds if a match could be established for all the words in a gloss term, and the sentiment scores become the average sentiment scores of the matching synset terms. The relaxed condition on multi-word terms solves additional 7.9% of the cases. Finally, if no match could be established across all the different gloss terms (1.2% of the entries), default neutral sentiment scores are assigned. The analysis of such cases is discussed in section 4.

Sometimes, a multi-word gloss term consists of words that denote excess (e.g., *most* and *more*), scarcity (e.g., *less* and *few*) or negation (e.g., *not*). We do not match such words to synset terms. Instead, they affect the polarity scores; we double the score, halve the score and swap the sign, respectively. We created the list of such words manually by examining AraMorph glosses.

Figure 3 illustrates the linking process between a sample of the processed AraMorph with a sample of the processed SentiWordNet, resulting in the construction of SLSA. The final SLSA lexicon consists of 34,821 entries. The counts of the different POS tags in SLSA along with the percentages of the different sentiment classes are reported in Table 1, while examples from the final lexicon are listed in Table 2.

POS	Count	Neutral %	+ve %	-ve %	Mixed %
NOUN	20,263	58.1	12.7	15.4	13.8
VERB	9,117	42.5	19.3	21.9	16.4
ADJ	5,395	36.5	18.7	17.0	27.8
ADV	46	73.9	10.9	2.2	13.0
ALL	34,821	50.7	15.3	17.3	16.7

Table 1: Statistics of SLSA: The counts of the different POS tags and the percentages of the different sentiment classes.

Lemma	POS	English Gloss
baliy~ap_1	NOUN	affliction/tribulation
balA_1	VERB	afflict/test

Processed AraMorph

Synset Term	POS	+ve	-ve
affliction	NOUN	0	0.625
afflict	VERB	0.0625	0.75

Processed SentiWordNet

Lemma	POS	English Gloss	+ve	-ve	Objective
baliy~ap_1	NOUN	affliction/tribulation	0	0.625	0.375
balA_1	VERB	afflict/test	0.0625	0.75	0.1875

SLSA

Figure 3: The linking between SentiWordNet and AraMorph by matching the AraMorph normalized glosses to the synset terms in SentiWordNet with respect to POS. The upper two tables are samples of the processed AraMorph and SentiWordNet, respectively, while the lower table represents a sample of SLSA based on the linking process. The objective score is calculated as *1 -(positive score+negative score)*.

Lemma	POS	English Gloss	+ve	-ve	Obj.
niEom_1	NOUN	wonderful	0.8	0	0.2
tawaE~aY_1	VERB	be attentive/cautious	0.4	0	0.6
AiHotiyAj_1	NOUN	need;requirement	0.1	0.2	0.7
$ahoriy~_1	ADJ	monthly	0	0	1
katab_1	VERB	write	0	0	1
mulaT~ax_1	ADJ	stained/sullied	0	0.3	0.7
dana>_1	VERB	be vile;be despicable	0	0.5	0.5
kamod_1	NOUN	swarthiness;sadness	0	0.8	0.2

Table 2: Examples of SLSA entries; Obj. = Objective. All scores are rounded for readability.

4 Evaluation

4.1 Intrinsic Evaluation

As mentioned in section 3, no match could be established for 1.2% of AraMorph entries. We manually investigate these cases more closely. About 75% of the entries that are not covered in SLSA have lemmas that express Arabic or Islamic subjects that do not have English counterparts such as *hamozap* (an Arabic name) and *kunAfap* (an Arabic food). Another 5% of the cases are countries or nationalities that are not listed in SentiWordNet such as *EAjiy* (Ivorian). Additional 2% of the cases are due to misspelled or non-English glosses in AraMorph such as *bon appetit*. The remaining cases (around 18%) have glosses that do not match any of the synset terms in SentiWordNet.

We then conduct an intrinsic evaluation of SLSA where the performance is compared to that of ArSenL, which is the most similar state-of-the-art lexicon (see Section 2). First, we randomly select 400 (lemma, POS) pairs for the evaluation. Only four pairs (1%) are not covered in SLSA. On the other side, 103 pairs (26%) are absent in ArSenL, which is consistent with the claim of the authors of ArSenL that only 76% of SAMA entries are matched in SentiWordNet. We then evaluate the random entries that exist in both SLSA and ArSenL (297 entries). We ask human anno-

tators to judge the correctness of the values in the two lexicons. ArSenL may have several sentiment values for the same entry, each with its own confidence score, so we used the sentiment values with the highest confidence score (averaged in the case of multiple answers). Since judging the values as real numbers is hard for humans, we map the sentiment scores into three classes of intensity (zero, up to 0.55 and above 0.55). An entry is correct only if the values of the positive and negative polarity classes are both correct. Each entry was judged by two annotators (without knowing its origin). They had to discuss and come to an agreement in the cases of disagreement (about 15% of the cases). SLSA and ArSenL have the exact same scores in 58.2% of the cases, which increases to 83.5% when mapping to the intensity classes.

Table 3 lists the accuracy of a majority baseline (neutral), SLSA and ArSenL for the different POS types[2]. SLSA gives error reductions of 58.7% and 37.8% over the baseline and ArSenL, respectively.

About 93% of SLSA errors are cases where the sentiment scores are doubtful in SentiWordNet, while the other errors are due to incorrect glosses in AraMorph. It might happen that an AraMorph entry is incorrectly linked to a SentiWordNet entry causing an error, but we do not see this in any of the manually analyzed data.

POS	Count	Baseline %	ArSenL %	SLSA %
NOUN	183	57.4	71.6	81.4
ADJ	50	42.0	66.0	74.0
VERB	62	43.5	58.1	80.6
ADV	2	50.0	100.0	100.0
All	293	51.9	68.0	80.1

Table 3: Accuracy results of a majority baseline (neutral), SLSA and ArSenL, evaluated on a test set that is covered in both SLSA and ArSenL

4.2 Extrinsic Evaluation

We conduct an extrinsic evaluation of SLSA on the task of sentiment analysis where a subjective sentence is classified to be either positive or negative. The performance is compared to that of ArSenL. We use an evaluation setup similar to the one described in (Badaro et al., 2014) using the corpus developed by Abdul-Mageed et al. (2011). The corpus involves 400 documents from the Penn Arabic Treebank (part 1 version 3) (Maamouri et al., 2004) where the sentences are tagged as objective, subjective-positive, subjective-negative and subjective-neutral. The evaluation only involves the sentences tagged as subjective-positive

and subjective-negative. Random 80% of the sentences are used for training, while the rest are left for testing.

We train a Support Vector Machines classifier, through LIBSVM (Chang and Lin, 2011), using sentence vectors of three features representing the averages of the positive scores, negative scores and objective scores of the non-stop words in the sentence divided by the count of the underlying words. The scores are obtained by querying the lexicon using the lemma and POS information.

We optimize the classification to obtain the best F1-score based on five-fold cross validation on the training set using different SVM kernels and parameters. Polynomial kernels give the best weighted-average F1-score[3] of 68.6% (using SLSA), which is an absolute 0.2% improvement over linear kernels. Table 3 lists the precision, recall and F1-score of a majority baseline (subjective-negative), SLSA and ArSenL. SLSA provides absolute weighted-average F1-score improvements of 22.9% and 3.5% over the baseline and ArSenL, respectively.

		Baseline %	ArSenL %	SLSA %
Positive	F1	0.0	54.5	58.5
	Precision	0.0	56.5	61.8
	Recall	0.0	52.5	55.6
Negative	F1	75.4	72.0	75.2
	Precision	60.6	70.4	72.8
	Recall	100.0	73.7	77.6
Weighted-Ave. F1		45.7	65.1	68.6

Table 4: Sentiment analysis results of a majority baseline (subjective-negative), SLSA and ArSenL

5 Conclusion and Future Work

We have presented a publicly available large-scale Standard Arabic Sentiment Lexicon (SLSA) that avoids the deficiencies in the current lexicons. The construction of SLSA is based on linking the lexicon of AraMorph with SentiWordNet along with a few heuristics and powerful back-off. SLSA has the highest up-to-date reported coverage. SLSA shows a relative improvement of 37.8% over a state-of-the-art lexicon when tested for accuracy. It also outperforms it by an absolute 3.5% of F1-score when tested for sentiment analysis.

The future plans include manually correcting SLSA to reach a nearly 100% accuracy. Additionally, the work will be extended to the Arabic dialects for which AraMorph-like morphological analyzers are available. We also plan to study the cases where English and Arabic translations have different sentiments due to cultural differences.

---

[2]There are only few adverbs in the test set because they are rare in Arabic, where only 0.1% of the lexicon entries are adverbs.

[3]The weighted-average F1-score is the sum of the F-1 score of the positive class and the F1-score of the negative class, each multiplied by its percentage.

References

Muhammad Abdul-Mageed and Mona Diab. 2012. Toward building a large-scale Arabic Sentiment Lexicon. In *The Sixth International Global Word-Net Conference*, Matsue, Japan.

Muhammad Abdul-Mageed and Mona Diab. 2014. SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis. In *The Ninth international conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland.

Muhammad Abdul-Mageed, Mona Diab, and Mohammed Korayem. 2011. Subjectivity and Sentiment Analysis of Modern Standard Arabic. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Oregon, USA.

Samah Alhazmi, William Black, and John McNaught. 2013. SentiWordNet in Relation to SentiWordNet 3.0. *International Journal of Computational Linguistics*, 4.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *The Seventh international conference on Language Resources and Evaluation (LREC)*, Malta.

Gilbert Badaro, Ramy Baly, Hazem Hajj, Nizar Habash, and Wassim El-Hajj. 2014. A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining. In *Arabic Natural Language Processing Workshop (WANLP), EMNLP*, Doha, Qatar.

William Black, Sabri Elkateb, Horacio Rodriguez, Musa Alkhalifa, Piek Vossen, Adam Pease, and Christiane Fellbaum. 2006. Introducing the Arabic WordNet Project. In *The Third International WordNet Conference (GWC)*, Jeju Island, Korea.

Tim Buckwalter. 2004. *Buckwalter Arabic Morphological Analyzer Version 2.0*. Linguistic Data Consortium, Philadelphia.

Chih-Chung Chang and Chih-Jen Lin. 2011. Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2.

Mona Diab, Mohamed Al-Badrashiny, Maryam Aminian, Mohammed Attia, Pradeep Dasigi, Heba Elfardy, Ramy Eskander, Nizar Habash, Abdelati Hawwari, and Wael Salloum. 2014. Tharwa: A Large Scale Dialectal Arabic - Standard Arabic - English Lexicon. In *The Ninth international conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland.

Alaa El-Halees. 2011. Arabic Opinion Mining Using Combined Classification Approach. In *The International Arab Conference on Information Technology (ACIT)*, Amman, Jordan.

Mohamed Elarnaoty, Samir AbdelRahman, and Aly Fahmy. 2012. A Machine Learning Approach For Opinion Holder Extraction in Arabic Language. In *Corr*, Amman, Jordan.

Mohamed Elhawary and Mohamed Elfeky. 2010. Mining Arabic Business Reviews. In *IEEE International Conference on Data Mining Workshops*, Sydney, Australia.

David Graff, Mohamed Maamouri, Basma Bouziri, Sondos Krouna, Seth Kulick, and Tim Buckwalter. 2009. Standard Arabic Morphological Analyzer (SAMA) Version 3.1. Linguistic Data Consortium LDC2009E73.

Mohamed Maamouri, Ann Bies, Tim Buckwalter, and Wigdan Mekki. 2004. The penn arabic treebank: Building a large-scale annotated arabic corpus. In *NEMLAR Conference on Arabic Language Resources and Tools*, Cairo, Egypt.

Fawaz H.H. Mahyoub, Muazzam A. Siddiqui, and Mohamed Y. Dahab. 2014. Building an Arabic Sentiment Lexicon Using Semi-supervised Learning. *Journal of King Saud University - Computer and Information Sciences*, 26.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *The 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Sydney, Australia.

George A. Miller, Richard Beckwith, and Derek Gross Christiane Fellbaum, and Katherine Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3.

Hanaa Mobarz, Mohsen Rashwan, and Ibrahim Abdel-Rahman. 2011. Generating lexical Resources for Opinion Mining in Arabic Language Automatically. In *The Eleventh Conference on Language Engineering (ESOLE)*, Cairo-Egypt.

Mike Thelwall, Kevan Buckley, Georgios Paltoglou, and Di Cai. 2010. Sentiment Strength Detection in Short Informal Text. *Journal of the American Society for Information Science and Technology*, 61.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *The First conference on Human Language Technology and Empirical Methods in Natural Language Processing (EMNLP)*, Vancouver, Canada.