A Large Scale Arabic Sentiment Lexicon for Arabic Opinion Mining

Gilbert Badaro, Ramy Baly, Hazem Hajj
Electrical and Computer Engineering Department
American University of Beirut, Lebanon
{ggb05,rgb15,hh63}@aub.edu.lb

Nizar Habash
Computer Science Department
New York University Abu Dhabi, UAE
nizar.habash@nyu.edu

Wassim El-Hajj
Computer Science Department
American University of Beirut, Lebanon
we07@aub.edu.lb

Abstract

Most opinion mining methods in English rely successfully on sentiment lexicons, such as English SentiWordnet (ESWN). While there have been efforts towards building Arabic sentiment lexicons, they suffer from many deficiencies: limited size, unclear usability plan given Arabic's rich morphology, or non-availability publicly. In this paper, we address all of these issues and produce the first publicly available large scale Standard Arabic sentiment lexicon (ArSenL) using a combination of existing resources: ESWN, Arabic WordNet, and the Standard Arabic Morphological Analyzer (SAMA). We compare and combine two methods of constructing this lexicon with an eye on insights for Arabic dialects and other low resource languages. We also present an extrinsic evaluation in terms of subjectivity and sentiment analysis.

1 Introduction

Opinion mining refers to the extraction of subjectivity and polarity from text (Pang and Lee, 2005). With the growing availability and popularity of opinion rich resources such as online review sites and personal blogs, opinion mining is capturing the interest of many researchers due to its significant role in helping people make their decisions (Taboada et al., 2011). Some opinion mining methods in English rely on the English lexicon SentiWordnet (ESWN) (Esuli and Sebastiani, 2006; Baccianella et al., 2010) for extracting word-level sentiment polarity. Some researchers used the stored positive or negative connotation of the words to combine them and derive the polarity of the text (Esuli and Sebastiani, 2005).

Recently, special interest has been given to mining opinion from Arabic texts, and as a result, there has also been interest in developing an Arabic Lexicon for word-level sentiment evaluation. The availability of a large scale Arabic based SWN is still limited (Alhazmi et al., 2013; Abdul-Mageed and Diab, 2012; Elarnaoty et al., 2012). In fact, there is no publicly available large scale Arabic sentiment lexicon similar to ESWN. Additionally there are limitations with existing Arabic lexicons including deficiency in covering the correct number and type of lemmas.

In this paper, we propose to address these challenges, and create a large-scale sentiment lexicon benefiting from available Arabic lexica. We compare two methods with an eye towards creating such resources for other Arabic dialects and low resource languages. One lexicon is created by matching Arabic WordNet (AWN) (Black et al., 2006) to ESWN. This path relies on the existence of a wordnet, a rather expensive resource; while the second lexicon is developed by matching lemmas in the SAMA (Graff et al., 2009) lexicon to ESWN directly. This path relies on the existence of a mere dictionary, still expensive but more likely available than a wordnet. Finally, the combination of the two lexicons is used to create the proposed large-scale Arabic Sentiment Lexicon (ArSenL). Each lemma entry in the lexicon has three scores associated with the level of matching for each of the three sentiment labels: positive, negative, and objective.

The paper is organized as follows. A literature review presented in section 2 is conducted on work that involved developing multilingual lexi-

cal resources. In section 3, the steps followed to create ArSenL are detailed. Extrinsic evaluation of ArSenL is discussed in section 4. In section 5, we conclude our work and outline possible extensions.

2 Literature Review

There have been numerous efforts for creating sentiment lexica in English and Arabic. Esuli and Sebastiani (2006) introduced English SentiWordNet (ESWN), a resource that associates synsets in the English WordNet (EWN) with scores for objectivity, positivity, and negativity. ESWN has been widely used for opinion mining in English (Denecke, 2008; Ohana and Tierney, 2009). Staiano and Guerini (2014) introduced DepecheMood, a 37K entry lexicon assigning emotion scores to words. This lexicon was created automatically by harvesting social media data and affective annotated data.

In the context of developing sentiment lexica and resources for Arabic, Abdul-Mageed et al. (2011) evaluated the use of an adjective polarity lexicon on a manually annotated portion of the Penn Arabic Treebank. They describe the process of creating the adjective polarity lexicon (named SIFAAT) in Abdul-Mageed and Diab (2012) using a combination of manual and automatic annotations. The manual annotation consisted of extracting 3,982 Arabic adjectives from the Penn Arabic Tree (part 1) and manually labeling them into three tags: positive, negative or neutral. The automated annotation relied on the automatic translation of the ESWN synsets and glosses using Google translate. More recently, Abdul-Mageed and Diab (2014) extended their lexicons creating SANA, a subjectivity and sentiment lexicon for Arabic. SANA combines different pre-existing lexica and involves extensive manual annotation, automatic machine translation and statistical formulation based on pointwise mutual information. The process also involved gloss matching across several resources such as THARWA (Diab et al., 2014) and SAMA (Graff et al., 2009). SANA included 224,564 entries which cover Modern Standard Arabic (MSA) as well as Egyptian and Levantine dialects. These entries are not distinct and possess many duplicates. Through these different publications, the authors heavily rely on two types of techniques: manual annotations, which can be rather expensive (yet accurate) and automatic translation which is cheap (but very noisy since the Arabic output is not diacritized and no POS information was used). Their SANA lexicon has a mix of lemmas and inflected forms, many of which are not diacritized. This is not a problem in itself, but it limits the usability of the resource. That said, we use their annotated PATB corpus and SIFAAT lexicon for evaluating our lexicon. We focus on these two resources because they were manually created and are of good quality.

Alhazmi et al. (2013) linked the Arabic WordNet to ESWN through the provided synset offset information. Their approach had limited coverage (~10K lemmas only) and did not define a process for using the lexicon in practical application given Arabic's complex morphology. Furthermore it is not yet publicly available and was not evaluated in the context of an application.

In addition to English and Arabic sentiment lexica development, recent efforts were put to develop a multilingual sentiment lexicon. Chen and Skienna (2014) proposed an automatic approach for creating sentiment lexicons for 136 major languages that include Arabic by integrating several resources to create a graph across words in different languages. The resources used were Wiktionary, Machine translation (Google), Transliteration and WordNet. They created links across 100,000 words by retrieving five binary fields using the above four resources. Then using a seed list obtained from Liu's English lexicon (2010) the sentiment labels are propagated based on the links in the developed graph. The resulting Arabic sentiment lexicon which is of small size was compared to SIFAAT (Abdul-Mageed and Diab, 2012).

We are inspired by these efforts for Arabic sentiment lexicon creation. We extend them by comparing different methods for creating such a resource with implications for other languages. Our lexicon is not only large-scale with high coverage and high accuracy, but it is also publicly available. Finally, our lemma-based lexicon is linked to a morphological analyzer for ease of use in conjunction with Arabic lemmatizer such as MADA (Habash and Rambow, 2005).

3 Approaches to Lexicon Creation

We define our target Arabic Sentiment Lexicon (or ArSenL) as a resource, pairing Arabic lemmas used in the morphological analyzer SAMA with sentiment scores such as those used in ESWN (positive, negative and neutral scores). We briefly describe next the different resources we use, followed by two methods for creating ArSenL: using an existing Arabic WordNet or using English glosses in a dictionary.

3.1 Resources

We rely on four existing resources to create ArSenL: English WordNet (EWN), Arabic WordNet (AWN), English SentiWordNet (ESWN) and SAMA. A high level summary of characteristics is shown in Table 1.

Lexicon	Language	Sentiment	#Synsets	#Lemmas
EWN	English	No	~90K	~120K
AWN	Arabic	No	~10K	~7K
ESWN	English	**Yes**	~90K	~120K
SAMA	Arabic-English	No	N/A	~40K
ArSenL	**Arabic**	**Yes**	157,969	**28,760**

Table 1. The different resources used to build ArSenL.

The English WordNet (EWN) (Miller et al., 1990) is perhaps one of the most used resources for English NLP. Several offset-linked versions of EWN have been released (2.0, 2.1, 3.0 and 3.1). The offset is a unique identifier for a synset in EWN. EWN includes a dictionary augmented with lexical relations (synonymy, antonymy, etc.) and part-of-speech (POS) tags.

Arabic WordNet (AWN 2.0) (Black et al., 2006) was part of a Global WordNet project whose aim was to develop WordNets similar to EWN but for different languages. AWN entries are connected by offsets to EWN 2.0. AWN does not include Arabic examples or glosses as EWN, but include POS tags.

English SentiWordNet (ESWN 3.0) (Esuli and Sebastiani, 2006) is a large-scale English Sentiment lexicon that provides for each synset in EWN 3.0 three sentiment scores whose sum is equal to 1: Pos, Neg, and Obj. ESWN has the same offset mappings of EWN across its different versions.

Standard Arabic Morphological Analyzer (SAMA 3.1) (Graff et al., 2009) is a commonly used morphological analyzer for Arabic. Each lemma has a POS tag and English gloss. The analyzer produces for a given word all of its possible readings out of context.

3.2 Arabic WordNet-based Approach

In this approach, we rely on the existence of a richly annotated resource, namely a wordnet, which is aligned to the ESWN. For Arabic, this approach requires two steps: mapping AWN to ESWN and mapping SAMA to AWN. The mapping between AWN to EWSN provides us with the sentiment scores and the mapping between AWN and SAMA provides us with the correct lemma forms for the words in AWN. We refer to the resulting lexicon as **ArSenL-AWN**.

Mapping AWN to ESWN. The entries in the various Wordnet resources we use are nicely linked through offsets to allow backward compatibility and linkage (see Figure 1). Figure 1 shows the connection with a walking example for the word شَعْر **$aEor**[1] **'**hair'. We use the available offset maps to link synsets in AWN 2.0 to those in ESWN 3.0 and thus are able to assign sentiment scores to the AWN 2.0 entries. We make use of sense map files provided by WordNet that connect its three different versions 2.0, 2.1 and 3.0. Since some of the offsets were used to refer to different entries in WordNet, POS tags were also checked to validate the mapping. The process of aligning AWN to ESWN yielded very reliable links.

We manually checked each of the 9,692 terms in AWN and their ESWN English complements. Out of the 9,692, there were only 9 AWN words that did not match with anything in ESWN; and 48 entries in AWN that had no lemmas to start with although they were linked to ESWN. These terms were dropped for the next processing performed. Thus, this technique only allowed us to line 9,635 synsets corresponding to 6,967 Arabic lemmas. Through this process, we noticed that there were no sense map files for adjectives in WordNet which limited the mappings performed in this approach to nouns and verbs only.

Mapping SAMA to AWN. The alignment of Arabic lemmas in SAMA and AWN is complicated due to several issues:

a. SAMA and AWN do not always agree on lemma orthography, e.g., long vowel A is represented as A in SAMA and aA in AWN, and the two resources do not always agree on Hamzated Alif forms (Habash, 2010). The issue of Hamzated Alif is solved by replacing it in both resources by the letter A. The definition of lemmas varies between the two, e.g., SAMA does not use the definite article in nouns, and uses the stem of the 3rd person masculine singular verb (as opposed to full form): katab not kataba 'to write'.
b. AWN has multi-word lemmas, which SAMA lacks.

[1] Arabic transliteration is provided in the Buckwalter Scheme (Buckwalter, 2004).

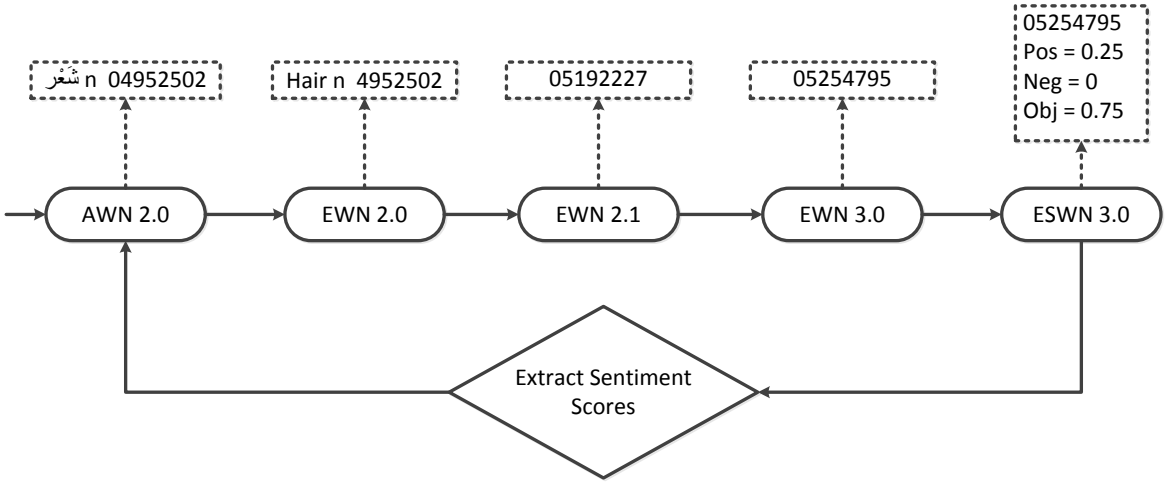

Figure 1. Steps to map AWN 2.0 to ESWN 3.0 with a walking example for the word شَعْر $aEor 'hair'.

To address these issues, we first exclude the multi-word lemmas in AWN, which account for 1,695 lemmas out of 6,967 (24%). Of the rest, exact matching against SAMA yields pairings for 1,736 lemmas. After applying a set of orthographic and lemma-form normalization rules as indicated in Table 2, exact matching yields additional 1927 lemma matches. Finally, we back off to using the SAMA morphological analyzer on AWN terms and selecting the lemma with the lowest edit distance. This step adds 1,094 lemma matches. Overall, 7,326 synsets entries corresponding to 5,002 lemmas in AWN are linked to 4,507 lemmas in SAMA. The linked lemmas account for 95% of all single word lemmas in AWN, but only correspond to 12% of SAMA lemmas. Moreover, we manually validated the mapping between SAMA and AWN lemmas, specifically the ones that were mapped using SAMA back off with minimum edit distance computation. 10% were not correct matches. We corrected them and created a gold reference for the lexicon, which we use in the evaluation section. In Table 3, we report some entries that were mapped wrongly between AWN and SAMA and which were removed from the lexicon.

In AWN	After Modification	Example
aA	A	(struggle) kifaAH → kifAH
If (pos = = v and lemma ends with a)	Remove "a"	(circulate) $aAEa →$aAE
If lemma ends with K	Replace K by iy	(past) mADK → mADiy

Table 2. Summary of modifications performed to AWN lemmas in order to match them to SAMA.

Examples of entries in ArSenL-AWN are shown in Table 4. Each row represents a field in ArSenL-AWN. AWN-Offset represents the offset of the Arabic word in AWN 2.0. SWN-Offset is the mapped SWN 3.0 entry's offset. The AWN lemma is the lemma form that is found in AWN 2.0 and SAMA lemma field is its corresponding lemma in SAMA form after performing the cleanup. Positive and negative score fields are the ones retrieved from SWN 3.0. The confidence is a percentage that represents our confidence in the entry.

AWN Offset	114276721	112853471	200548789
SWN Offset	15133621	13619764	00564300
POS tag	N	n	v
AWN Lemma	>amad_n1AR	AlgaAluwn_n1AR	Haloma>a_v1AR
SAMA Lemma	>amobiyr_1	gAliy_1	Halum-u_1
Positive Score	0	0	0
Negative Score	0	0	0
Confidence	100	100	100
English Gloss	Duration	gallon	hydrolize

Table 3. Examples of entries that were mapped incorrectly from AWN to SAMA

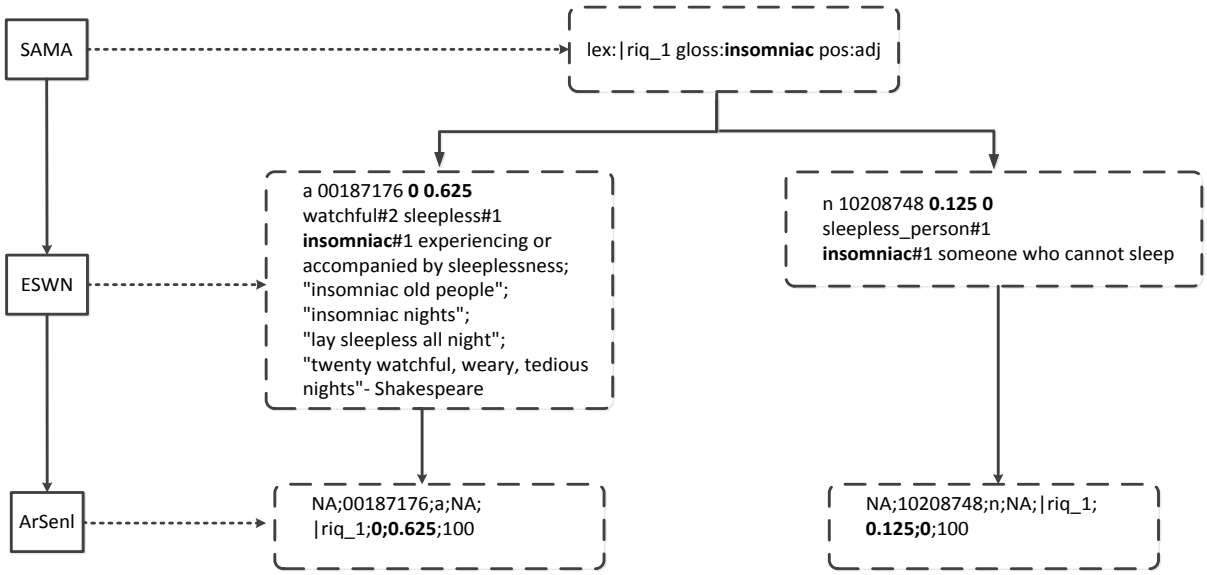

Figure 2. Steps to map SAMA to ESWN 3.0 with a walking example for the word أرق.

Since AWN was connected to SWN through a direct mapping all the entries of ArSenL-AWN were assigned 100% confidence. In table 5, row 3 summarizes the numbers obtained through the automated process and row 7, the results obtained after manual correction.

AWN Offset	100392523	201014980
SWN Offset	00410247	01048569
POS tag	n	v
AWN Lemma	EaAdap_n1AR	SaAHa_v2AR
SAMA Lemma	EAdap_1	SAH-i_1
Positive Score	0.25	0
Negative Score	0.125	0
Confidence	100	100
English Gloss	habit, custom, practice	scream, call

Table 4. Examples of entries in ArSenL-AWN.

3.3 English Gloss-based Approach

In this approach, we make use of the English glosses associated with the SAMA lemma entries. For each entry, we find the synset in ESWN with the highest overlap in SAMA English glosses. A walking example of the described method is shown in Figure 2. The recall of the SAMA gloss is used as a confidence measure of the mapping. We refer to the resulting lexicon as **ArSenL-Eng**.

Each lemma in SAMA is appended with a gloss list that varies in size from 1 up to 6 words. Let *n* denote the number of words available in the gloss list. We first attempt to match all the words in the list to the glosses of each entry in ESWN. If one or more matches are found, the scores are retrieved and a new entry in SAMA is processed as described. In case there were no matches, we try to find an overlap between a combination of *n-1* words of the SAMA gloss list and the glosses of ESWN. If one or more matches are found, the scores are retrieved and each match is recorded in ArSenL-Eng. Again, if no matches were obtained, the same process is repeated for the combination of *n-2* words of the SAMA gloss list.

Lexicon	#Lemmas	#Related Synsets
Automatic Process		
ArSenL-AWN	4,507	7,326
ArSenL-Eng	28,540	150,700
ArSenL-Union	28,812	158,026
Manual Correction		
ArSenL-AWN	4,492	7,269
ArSenL-Union	28,780	157,969

Table 5. Sizes of the created sentiment lexica.

This procedure is followed until we span all the words in the gloss list. As the number of words used in the combination to check for overlap between the two resources decreases, the confidence percentage decreases. In ArSenL-Eng, the confidence measure is equal to the ratio of the number of words overlapping between SAMA and ESWN over the total number of words available in the gloss list of the corresponding SAMA entry. Besides checking the overlap of glosses, POS tags are also used to make sure that verbs are not mapped to nouns and vice versa. This technique results in mapping 150,700 ESWN

synsets corresponding to 28,540 distinct lemmas in SAMA (76%). The validation of ArSenL-Eng was performed (a) automatically by using **ArSenL-AWN** and (b) manually by randomly validating 400 distinct lemmas. For the automated part, we check for each common lemma between the two lexicons if the sentiment scores match. A total of 3,833 lemmas (out of 4,507) from ArSenL-AWN were matched in ArSenL-Eng.

Thus, we can inspect that the precision of the remaining scores is of 85%. For the manual validation, we check if the meaning of the SAMA lemma corresponds to the one in ESWN. 70% of the 400 randomly selected lemmas were accurately mapped to ESWN. The main issue of the remaining 30% is the unavailability of enough glosses per SAMA lemma, which makes the connection weaker. This approach did not involve manual correction and the lemma numbers are reported in row 4 of Table 5 along with their corresponding number of related synsets.

3.4 Combining the Two Approaches

We combine the two lexica created above by taking their union. We refer to the resulting lexicon as ArSenL. The details of the sizes of the three lexica are shown in Table 5.

The union of the two lexicons consisted of combining the two resources and adding a field in the lexicon to distinguish the original source of the entry. For instance, an entry from the first approach, i.e. ArSenL-AWN, will have an AWN offset while an entry in ArSenL-Eng will have the same field set to N.A (Not Available). Furthermore, due to manual correction performed to ArSenL-AWN, the gold version of the union lexicon includes 28,780 lemmas with the corresponding number of 157,969 synsets.

A public interface to browsing ArSenL is available at http://oma-project.com. The interface allows the user to search for an Arabic word. The output would show the different scores for the Arabic word along with the corresponding sentiment scores, English glosses and examples that help in disambiguating different sentiment scores for the same Arabic lemma. Work is also being done to allow searching for English words and finding the corresponding Arabic words. Snapshot of the homepage is shown in Figure 3.

4 Evaluation

We conduct an extrinsic evaluation to compare the different versions of ArSenL on the task of subjectivity and sentiment analysis (SSA). We also compare the performance of the SIFAAT lexicon (Abdul-Mageed et al., 2011) discussed in Section 2.

Experimental Settings We perform our experiments on the same corpus used by Abdul-Mageed et al. (2011). The corpus consists of 400 documents form the Penn Arabic Treebank (part 1 version 3) that are gold segmented and lemmatized. The sentences are tagged as objective, subjective-positive, subjective-negative and subjective-neutral.

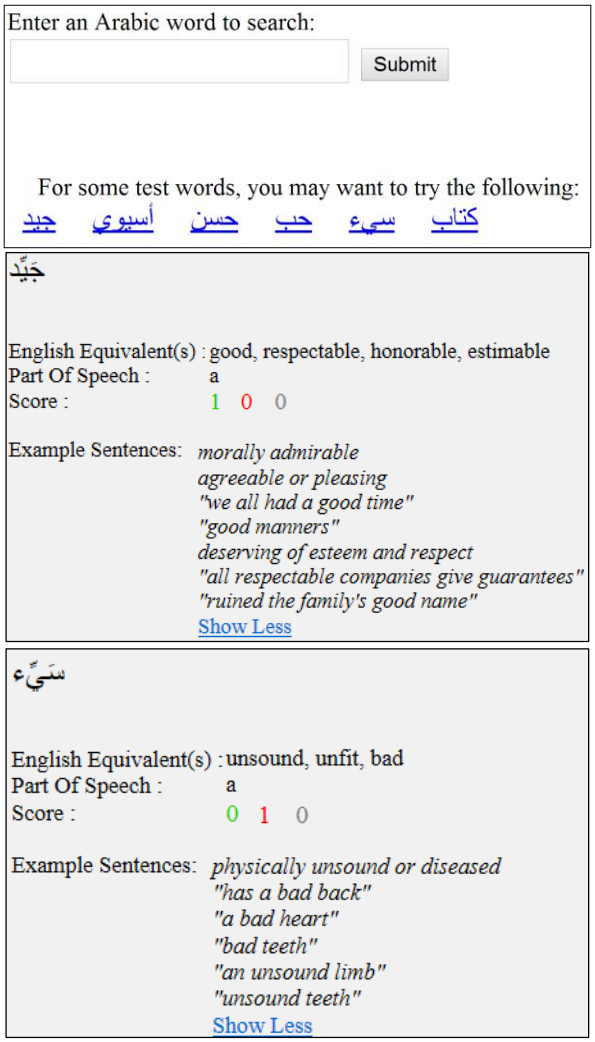

Figure 3. Homepage of the lexicon interface and snapshots of examples searched through the interface. Positive, negative and objective scores are represented in green, red and gray respectively.

We use nonlinear SVM implementation in MATLAB, with the radial basis function (RBF) kernel, to evaluate the different lexicons in the context of SSA. The classification model is developed in two steps. In the first step, the kernel parameters (kernel's width γ and regularization parameter C) are selected, and in the second step the classification model is developed and evalu-

ated based on the selected parameters. To decide on the choice of RBF kernel parameters, we use the first 80% of the dataset to tune the kernel parameters to the values that produce the best F1-score using 5-fold cross-validation. The resulting parameters are then used to develop and evaluate the SVM model using 5-fold cross-validation on the whole dataset.

Two experiments were conducted to evaluate the impact of the different lexicons on opinion mining. The first experiment considers subjectivity classification where sentences are classified as either subjective or objective. In this experiment, the SVM kernel parameters were tuned to maximize the F1-score for predicting subjective sentences. The second experiment considers sentiment classification, where only subjective sentences are classified as either positive or negative. Subjective-neutral sentences are ignored. In this experiment, the classifier's parameters are tuned to maximize the average F1-score of positive and negative labels. We report the performance measures of the individual classes, as well as their average.

For baseline comparison, the majority class is chosen in each of the experiments, where all sentences are assigned to the majority class. For subjective versus objective baseline classification, all sentences were classified as subjective since the majority (55.1%) of the sentences were subjective. To further emphasize the importance of detecting subjectivity, we chose the F1-score for subjective as baseline. For positive versus negative baseline classification, all sentences were classified as negative since the majority (58.4%) of the dataset was annotated as negative. The resulting baseline performance measures are captured in Table 6, and serve as basis for comparison with our developed models. For the subjective versus objective the baseline F1-score is 71.1%, and for positive versus negative, the baseline F1-score is averaged as 36.9%.

Features We train the SVM classifier using sentence vectors consisting of three numerical features that reflect the sentiments expressed in each sentence, namely positivity, negativity and objectivity. The value of each feature is calculated by matching the lemmas in each sentence to each of the lexicons separately: ArSenL-AWN, ArSenL-Eng, ArSenL-Union and SIFAAT. The corresponding scores are then accumulated and normalized by the length of the sentence. We remove all stop words in the process. For words that occur in the lexicon multiple times, the average sentiment score is used. It is worth noting that the choice of aggregation for the different scores and the choice of nonlinear SVM was concluded after a set of experiments, but not reported in the paper. In this regards, we conducted a suite of experiments to evaluate the impact of using: (a) linear versus Gaussian nonlinear SVM kernels, (b) normalization based on sentence length, (c) normalization using z-score versus not, and (d) using the confidence score from the lexicons. Our best results across the different configurations reflected the best results with the nonlinear Gaussian RBF kernels, with sentence length-based normalization and without confidence weighting.

		Baseline	ArSenL			Sifaat
			AWN	Eng	Union	
Coverage %		NA	56.6	88.8	**89.9**	32.1
Subjective	F1	71.1	71.2	72.1	**72.3**	66
	Pre	55.1	58.1	58.5	58.3	**61.5**
	Rec	100	92	93.9	**95.1**	71.4
Positive	F1	0	52.9	59.7	**61.6**	55.4
	Pre	0	44.7	55	**55.2**	51.8
	Rec	0	64.8	65.6	**70.1**	60.2
Negative	F1	73.7	55	65.1	**67.3**	63
	Pre	58.4	67	70.7	**75.6**	67.6
	Rec	100	46.9	60.6	**61**	59.4
Average F1 (Pos/Neg)		36.9	53.9	62.4	**64.5**	59.2

Table 6. Results of extrinsic evaluation. Numbers that are highlighted reflect the best performances obtained by the lexicons, without considering the baseline

Results Three evaluations were conducted to compare the performances of the developed sentiment lexicons. The results of the experiments are shown in Table 6. First, we evaluate the coverage of the different lexicons. We define coverage as the percentage of lemmas (excluding stop words) covered by each lexicon. ArSenL-AWN and SIFAAT have lower coverage than the ArSenL-Eng lexicon. The union lexicon has the highest coverage. This is normally due to the larger number of lemmas included in the English and union lexicons, as shown in Table 5.

In subjectivity classification, ArSenL lexicons perform better than the majority baseline and outperform SIFAAT in terms of F1-score. Overall, the developed ArSenL-Union gives the best performance among all lexicons. The only exception of better performance for SIFAAT for subjectivity is in terms of precision, which is associated with a much lower recall resulting in an F1-score that is lower than that of ArSenL's.

Similarly, sentiment classification experiment reveals that ArSenL lexicons produce results that are consistently better than SIFAAT and the majority baseline. The ArSenL-Union lexicon outperforms all lexicons in all measures without exceptions.

In summary, it can be observed that the English-based lexicon produces results that are superior to the AWN-based lexicon. Combining both resources, through the union, allows further improvement in SSA performance. It is also worth noting that the English and union lexicons consistently outperform SIFAAT despite the fact that the latter was manually derived from the same corpus we are using for evaluation. We close by showing examples of ArSenL in Table 7.

The lemmas are in their Buckwalter (2004) format for easier integration in any NLP task. The word NA stands for Not Applicable. In the case where AWN Offset is NA and AWN lemma is NA, this means that the entry is retrieved from ArSenL-Eng. Otherwise, the entries are from ArSenL-AWN. The additions to the lemmas such as "_v1AR" , "_n1AR", "_1" or "_2" can be dropped when data processing is performed. They were kept for easier retrieval in the original sources (AWN and SAMA). We added the "English Gloss" field for easier understanding of the Arabic word in the table. Moreover, it can be seen that only positive and negative scores are reported in the lexicon since the objective score can be easily derived by subtracting the sum of positive and negative scores from 1.

5 Conclusion and Future Work

We create a large sentiment lexicon for Arabic sentiments using different approaches linking to ESWN. We compared the two methods. Our results show that using English-based linking produces, on average, superior performance in comparison to using the WordNet-based approach. A union of the two resources is better than either and outperforms a high-quality manually-derived adjective sentiment lexicon for Arabic.

In the future, we plan to make use of this lexicon to develop more powerful SSA systems. We also plan to extend the effort to Arabic dialects and other languages.

6 Acknowledgments

This work was made possible by NPRP 6-716-1-138 grant from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors. Nizar Habash performed most of his contribution to this paper while he was at the Center for Computational Learning Systems at Columbia University.

AWN Offset	SWN Offset	POS tag	AWN Lemma	SAMA Lemma	Positive Score	Negative Score	Confidence	English Gloss
NA	04151581	n	NA	Aap_1	0	0	100	screen
NA	01335458	a	NA	$ATir_1	0.75	0	33	smart;bright
NA	05820620	n	NA	$Ahidap_1	0	0	50	proof
NA	00792921	v	NA	$Al-u_1	0	0	50	lift
NA	01285136	a	NA	$Amix_1	0.75	0	33	superior
NA	04730580	n	NA	danA'ap_1	0.222	0.778	33	inferiority
NA	01797347	v	NA	Hazin-a_1	0	0.5	50	sorrow
NA	00811421	a	NA	sAxin_1	0.75	0.125	50	hot
NA	07527352	n	NA	faraH_1	0.5	0.25	33	joy
NA	00064787	a	NA	Hasan_1	0.625	0	100	good
200300610	00310386	v	<izodahara_v1AR	{izodahar_1	0.125	0	100	flourish
200844607	00873682	v	>a$oEara_v1AR	>a$oEar_1	0	0	100	notify
201766276	01819147	v	>aHobaTa_v1AR	>aHobaT_1	0.125	0.5	100	discourage
114279405	15136453	n	nahaAr_n1AR	nahAr_2	0	0	100	day
100059106	00064504	n	najaAH_n1AR	najAH_2	0.625	0	100	success
113808178	14646610	n	naykl_n1AR	niykol_1	0	0	100	nickle
104540432	04748836	n	tabaAyun_n1AR	tabAyun_1	0.25	0.625	100	difference
200705236	00729378	v	tasaA'ala_v1AR	tasA'al_1	0.375	0	100	wonder
NA	01983162	a	NA	$ariyf_2	1	0	67	respectable
NA	05144663	n	NA	$ariyr_1	0	0.75	33	evil

Table 7. Samples of ArSenL showing entries originating from ArSenL-Eng and ArSenL-AWN.

References

Abdul-Mageed, M., Diab, M. and Korayem, M. (2011). Subjectivity and sentiment analysis of modern standard Arabic. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers*-Volume 2. Association for Computational Linguistics.

Abdul-Mageed, M., & Diab, M. (2012). Toward building a large-scale Arabic sentiment lexicon. In *Proceedings of the 6th International Global WordNet Conference* (pp. 18-22).

Abdul-Mageed, M., & Diab, M. (2014) SANA: A Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis. *In Proceedings of the Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland

Alhazmi, S., Black, W., & McNaught, J. (2013). Arabic SentiWordNet in Relation to SentiWordNet 3.0. *2180-1266*, *4*(1), 1-11.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *LREC* (Vol. 10, pp. 2200-2204).

Black, W., Elkateb, S., & Vossen, P. (2006). Introducing the Arabic wordnet project. In *In Proceedings of the third International WordNet Conference (GWC-06)*.

Buckwalter, T. 2004. Buckwalter Arabic morphological analyzer version 2.0. LDC catalog number LDC2004L02, ISBN 1-58563-324-0

Chen, Y., & Skienna, S. (2014). Building Sentiment Lexicons for All Major Languages. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)* (pp. 383-389). 2014 Association for Computational Linguistics.

Denecke, K. (2008, April). Using sentiwordnet for multilingual sentiment analysis. *In Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference* on (pp. 507-512). IEEE.

Diab, M., Al-Badrashiny, M., Aminian, M., Attia, M., Dasigi, P., Elfardy, H., Eskander, R., Habash, N., Hawwari, A., & Salloum, W. (2014). Tharwa: A Large Scale Dialectal Arabic - Standard Arabic - English Lexicon. *In Proceedings of the Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland.

Elarnaoty, M., AbdelRahman, S., & Fahmy, A. (2012). A Machine Learning Approach for Opinion Holder Extraction in Arabic Language. *International Journal of Artificial Intelligence & Applications*, *3*(2).

Esuli, A., & Sebastiani, F. (2005). Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM international conference on Information and knowledge management* (pp. 617-624). ACM.

Esuli, A., & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC* (Vol. 6, pp. 417-422).

Graff, D., Maamouri, M., Bouziri, B., Krouna, S., Kulick, S., & Buckwalter, T. (2009). Standard Arabic Morphological Analyzer (SAMA) Version 3.1, 2009. Linguistic Data Consortium LDC2009E73.

Habash, N., & Rambow, O., (2005). Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics.

Habash, N. (2010). Introduction to Arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, *3*(1), 1-187.

Liu, B. (2010). Sentiment Analysis and Subjectivity. *Handbook of Natural Language Proccessing*, 2:568.

Maamouri, M., Bies, A., Buckwalter, T., & Mekki, W. (2004, September). The penn arabic treebank: Building a large-scale annotated arabic corpus. In *NEMLAR Conference on Arabic Language Resources and Tools* (pp. 102-109).

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, *3*(4), 235-244.

Ohana, B., & Tierney, B. (2009, October). Sentiment classification of reviews using SentiWordNet. *In 9th. IT & T Conference* (p. 13).

Pang, B., & Lee, L. (2004, July). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics* (p. 271). Association for Computational Linguistics.

Pang, B., & Lee, L. (2005, June). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 115-124). Association for Computational Linguistics.

Staiano, J., & Guerini, M. (2014). DepecheMood: a Lexicon for Emotion Analysis from Crowd-Annotated News. *arXiv preprint arXiv:1405.1605*.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.