Automatic Identification of Rhetorical Elements in classical Arabic Poetry

Heyam Abd Alhadi <heyam_dot_abdalhade_at_gmail_dot_com>, University of Haifa
Ali Ahmad Hussein <ahussein_at_univ_dot_haifa_dot_ac_dot_il>, University of Haifa
Tsvi Kuflik <tsvikak_at_is_dot_haifa_dot_ac_dot_il>, University of Haifa

Abstract

A novel, rule-based, automatic framework for identifying rhetorical elements in classical Arabic poetry is described. Since rule-based approaches have well-known limitations, it is proposed as an interim solution until a sufficient quantity of annotated text has been amassed with which to train a machine-learning algorithm. The manual process of identifying rhetorical features in classical Arabic poetry is both time-consuming and requires high-level expertise in Arabic literature. Hence, automatic recognition system will solve this challenge. Automatic identification is, however, challenging, mainly because there is no existing annotated corpus with which to train a machine-learning-based classifier. The framework proposed here combines natural language-processing techniques with a rule-based reasoning approach, and will continually improve as more examples become available. It is intended as an initial step toward building the essential annotated corpus. Its focus is 20 rhetorical elements, all of importance according to classical Arabic rhetoricians, and it achieves the extremely encouraging result of an overall F-measure of 0.902.

1. Introduction

Rhetoric, *balāgha* in Arabic, is the art of persuasion as well as affectation through language. It is the means of conveying an idea in the most effective and convincing way, and with the most emotional impact [Moreau 2003]. In Arabic criticism, the term *balāgha* refers both to figures of speech (such as "simile") and stylistic — including embellishing — elements (such as "paronomasia") that are used by critics to analyse the aesthetic value of an artistic expression; mainly in classical Arabic poetry and the Qurʾān [Schaade and von Grunebaum 2012]. [1]

Since the ninth century CE, study of rhetoric has been stimulated by an increased interest in understanding the structure and development of poetry as well as by the need to rationalise the aesthetic implications of the theological postulate of the uniqueness (*iʿjāz*) of the Qurʾān [Schaade and von Grunebaum 2012]. [2]

Despite the great potential for study of rhetoric for so wide a range of research priorities both in and beyond itself, it remains lamentably neglected in modern research. Whereas there has been extensive research into Greek and mediaeval European rhetoric, there is nothing comparable for that of Arabic-Islamic texts, despite their high level of sophistication and their derivation from a major world civilisation. Among many factors which may explain this is the very large number of the rhetorical elements. In late rhetorical anthologies from the seventeenth-eighteenth century CE, their number is about one hundred and eighty. Identifying these elements in the thousands of classical Arabic poems that have been preserved from the pre-Islamic to the later Islamic periods, is a hard, if not impossible, mission. Tracing the development of rhetoric in this poetry is one way of understanding its development through ages. This is done by analysing their use in different poems during different periods, which is a lengthy and demanding undertaking far beyond a scope of a single study or researcher. [3]

In order to make this feasible, our Arabic Rhetoric Identifier (known as the REI) had to be developed.[1] The REI was created by a team led by A. Hussein and T. Kuflik. Its purpose is to enable researchers to annotate manually the rhetorical elements in Arabic poetry, and compile them into a database of rhetorical elements in classical Arabic poetry between the 5th and 16th centuries CE. The canon comprises 26,925 classical Arabic poems, only 75 of which have, to date, been manually analysed. The main challenges to researchers are the time-consuming nature of manual annotation and the high level of expertise required to identify and comprehend the properties of each element [see, for example, the elements in 7, 19, 20]. While automating this process is thus very necessary, it is also extremely challenging. [4]

While there are many modern studies of rhetorical usage in non-Arabic literatures, there is a lacuna in the modern era in Arabic literature (see, for example, Western literature's rich bibliography in comparison with its scarce listing for metaphor and metonymy in the *Bibliography of Metaphor and Metonymy* [2]). Scholars of classical Arabic poetry agree that the differing use of rhetoric is the main reason for the development of Arabic poetry in its different periods [16; cf. chapters 1, 2]. *Badīʿ* — literally *the new style* (or the novel, original, extraordinary style) — was the term used by Arabic rhetoricians to describe the sophisticated use of some rhetorical elements in this poetry, a use which renders the literary text differently sophisticated and differently aesthetic. In short, scholars agree that rhetoric was used in pre- and early Islamic poetry (5th-8th centuries CE), although it was considered more important in the ʿAbbāsid era (750–1517 CE). The *badīʿ* is the major element which distinguishes ʿAbbāsid from pre-ʿAbbāsid poetry — although which figurative speeches in *badīʿ* style and how they were used to make this difference remains unresolved. [5]

Despite the importance of *badīʿ*, studies which discuss it are few and incomplete ([Habash, Rambow, and Roth 2009], cf. pp. 28-41): [6]

1. Most discuss *badīʿ* in separate and randomly chosen verses rather than in complete poems. Considering this style in complete poems gives a far more accurate picture of the nature of *badīʿ* in classical Arabic poetry.
2. Since the main hypothesis of modern research (and of classical scholars) is that *badīʿ* is more important in *ʿAbbāsid* than in the pre-Islamic era, modern studies focus on the *badīʿ* of this poetry in the later period.

As far as we know, there are no comparative studies that analyze the development of *badīʿ* from one period to another, other than those recently conducted by one author of this article [Hussein 2015b], and by Wolfhart Heinrichs in 1994 [Heinrichs 1994]. Our approach is that the essence of *badīʿ* in classical Arabic periods, especially in ʿAbbāsid and pre-ʿAbbāsid times, cannot be described without identifying and analysing the rhetorical elements in complete poems. There are some 180 rhetorical elements described in classical books on rhetoric [Cachia 1998]. The number of classical Arabic poems, from the pre-Islamic era until the most recent pre-modern period, is enormous, making manual rhetorical analysis for the entire corpus a mission impossible. For this reason, Hussein and Kuflik began developing a system that will automatically analyse several rhetorical elements in an existing corpus of some 26,000 poems[3] (in development so access is currently restricted). This corpus includes poetry from pre-Islamic times to the Mamluk era, which ended early 16th century. All the poems were entered manually into the system by students of Arabic language and literature, according to printed versions of the poems. This article reports on the automatic analysis of 20 rhetorical elements. A manual analysis by A. Hussein of several poems in the present corpus showed these to be the main, recurrent rhetorical elements in the pre- and early Islamic period, including the ʿAbbāsid poetry in its first four centuries (that is, until the twelfth century CE). In addition, most of these elements (nos. 1, 2.3, 3.3, 10, 12, 14, 16, 17, 18, 19-20) are discussed in the first and oldest books on Arabic rhetoric, while the others (the rest of the 180 elements) appear only as later developments and feature in later rhetoric books. These were the two reasons for focusing on these elements in the present study. [7]

The classical automation approach uses text-mining and machine-learning techniques to identify rhetorical elements in the text. This approach requires collecting a large set of examples for each rhetorical element, and training a machine-learning-based classifier to recognize them. As we did not have an annotated set of examples, our first task was finding a way for assembling a sufficient number of analyzed poems to train a classifier. Our proposal aims at automating this step by proposing a framework to identify individual rhetorical elements in classical Arabic poetry by combining natural language processing (NLP) techniques with a rule-based approach. This framework has been evaluated on a small set of Arabic poems, and results have been extremely encouraging. [8]

The development of this automatic framework will make a vital contribution to the study of rhetoric in classical Arabic poetry. It will enable automatic recognition of the 20 selected rhetorical elements in a vastly larger corpus (26,925 poems) that cannot be handled manually. Analyzing the rhetoric in a corpus of this size will yield better understanding of the development of rhetoric in Arabic poetry from the pre-Islamic period to the 16th century CE. [9]

It is nonetheless important to note that rule-based systems have their own limitations [Hayes-Roth 1985], including the need for continuous maintenance and updating by experts – as demonstrated by this research. This is why we see it as an interim approach until sufficient sets of examples are available for training state-of-the-art machine learning algorithms for the task. [10]

Another challenge is the imprecise nature of such systems. Not only do they tend to misclassify examples, they also miss relevant elements (false negatives) and suggest wrong elements (false positives). Even the current framework, which achieved, overall, extremely encouraging results, had widely differing success rates for different elements. There is more work to do to ensure a success rate closer to 100%.

This paper is structured as follows. Section 2, which follows this introductory section, gives background and reviews related work. Section 3 describes the features of classical Arabic poems. Our framework is delineated in Section 4, together with the NLP tools and resources used in its implementation. The dataset, experiments and results are described in Section 5, and Section 6 discusses the study's implications and limitations. Conclusions and future work are considered in Section 7.

2. Background and Related Work

Over the years, considerable research efforts have been invested in developing methods and tools for NLP of Arabic. NLP of Arabic poetry, and, especially, identifying characteristics in this poetry, remain challenging.

Previous studies have focused on a range of aspects — morphology, morphological decomposition, morpho-syntactic analysis, sentiment analysis, semantic analysis and many others. One main research avenue is morphological analysis of Arabic texts, for which several tools exist. Among them are AlKhalil Morpho Sys 2 [Boudchiche et al. 2017], MADA [Diehl et al. 2009], MADA+TOKAN [Habash, Rambow, and Roth 2009], MADAMIRA [Pasha et al. 2014], ASMA [Abdul-Mageed, Diab, and Kübler 2013], MAGEAD [Habash and Rambow 2006], BAMA [Buckwalter 2002b], SAMA [Maamouri et al. 2010], SALMA [Sawalha, Atwell, and Abushariah 2013], Arabic Sentence-level [Shoukry 2013] and ArSenL [Badaro et al. 2014].

Of these, AlKhalil2 [Boudchiche et al. 2017] is the most recent and seems to be the best. It is an Arabic morpho-syntactic analyzer with a large database, which it uses to segment and then analyze partially or totally vocalized Arabic words. For each, the resulting analysis is derived as a set of features — prefix, stem, type, pattern of word, root, part of speech (POS), lemma, pattern of lemma and suffix. It outperforms BAMA and SAMA, achieving coverage of over 99%, as well as performing better in identifying various features, including those determined in the word context: the lemma, the stem and the diacritized form. This, therefore, was the tool used in our study.

ArSenL [Badaro et al. 2014] is a large sentiment lexicon for Arabic, which uses different approaches linked to the English SentiWordnet (ESWN). It has 154,396 entries, each referring to a single Arabic word, and each consisting of eight fields — Arabic morphologic lemma (1), POS (2), positive sentiment score (3), negative sentiment score (4), confidence (5), Arabic WordNet (AWN) offset (6), sentiment WordNet offset (7), and AWN lemma (8). An example is:

```
<imobiyriyAliy~_1; a;0;0.125;50; NIL;02746897; NIL
                 zrAEap_1; n;0;0;100;100861982;00916464; ziraAoEap_n1AR
```

> **Example 1.** Two entries in ArSenL. NIL = not in language

Given its large lexicon, it was selected for our study.

Another tool that we used was the Arabic WordNet (AWN) ([Fellbaum and Alkhalifa 2006], [Elkateb et al. 2006], [Habash, Rambow, and Roth 2009]) based on WordNet [Miller 1995], which are, respectively, lexical resources in standard Arabic and English. The Arabic WordNet is a large lexical database, comprising nouns, verbs, adjectives and adverbs, grouped in sets of cognitive synonyms, each expressing a distinct concept [Elkateb et al. 2006].

Hussein conducted pioneering manual studies to learn the use of rhetorical elements in classical Arabic poetry. These studies are few in number and deal with a very small corpus. In [Hussein 2018], he focuses on the use of a single rhetorical figure, the "intellectual trope" or *majāz ʿaqlī* (an important rhetorical element neglected in modern research) in one pre-Islamic wine poem. He explains in detail the idea of the *majāz ʿaqlī*, and discusses its semantic aspects in classical Arabic poetry, as shown in an incident concerning wine in a poem composed by Abū Dhuʾayb al-Hudhalī (d. ca. 648 CE). In another study [Hussein 2015a], he examines other rhetorical techniques and their use by Hudhalī poets in describing the element of wine. He suggests that rhetorical elements should be studied in as many classical Arabic poetry themes as possible. He also investigated [Hussein 2015b] rhetorical texture in two traditional tripartite poems — one dating from the pre-Islamic period, composed by ʿAlqama l-Faḥl (d. ca. 603 CE), the other from the early ʿAbbāsid era, composed by Bashshār b. Burd (d. 784 CE). The two have the same structure, address the same themes and comprise virtually the same number of verses. A key question concerns the rhetorical figures each poet uses, which leads to a broader question: How different was the *badīʿ* ("rhetorical") style in the poetry of the two eras, assuming that each poem is representative of its time?

Abuata and Al-Omari [Abuata and Al-Omari 2018] adopt the same rule-based approach as us for computerized analysis of Arabic prosody. They introduce an algorithm to determine the correct metre for a given Arabic poem and to convert the poem into metric writing.

A new way of recognizing the characteristics of classical Arabic poetry is presented by [Ahmed and Trausan-Matu 2017]. The authors focus on rhythm, which is considered the crown of this type of poetry. A stylistic study of Arab-American poetry [Pasha et al. 2014] attempts to analyze and describe the major features of this poetry, written by prominent Arab poets who immigrated to America during the 19th century. It examines their poetry according to linguistics, grammar, lexicon and rhetoric [Abushihab 2020].

A study of NLP for Arabic metaphors in [Alkhatib and Shaalan 2017] presents the challenges central to developing a computational NLP-based system for classical Arabic, Modern Standard Arabic and Dialect Arabic. It also highlights the main problems of translating Arabic metaphors into other languages.

This article proposes a method for supporting the study of rhetoric in Arabic literature. It demonstrates and evaluates an automatic rule-based framework for recognizing its 20 most important rhetorical elements, previously identified by experts in a time-consuming process. As we have seen, a wide variety of tools enables analysis of Arabic texts, and has paved the way for the present study — the first to examine automatic identification of rhetorical elements in classical Arabic poetry.

3. Features of classical Arabic Poems

Style is very much more salient than content in the historical development of the Arabic literary register, especially in classical Arabic poetry, where rhetoric is a primary feature. The development of rhetoric in literary texts, however, is significantly under-researched. Study of rhetoric and its development in practice are important not only for literary and critical studies, but also have considerable implications for our understanding of other fields of knowledge. The dense use of periphrasis in poems in one period, for example, followed by its relative displacement by simile in the next and by metaphor in the one after that may indicate and/or reflect cognitive-linguistic changes in Arab communities over time [Hussein n.d.].

Classical Arabic poetry is built of consecutive verses (lines in a poem), with each verse divided into two parts (hemistiches), the so-called *ṣadr* ‑صدر and *ʿajuz*‑عجز, respectively. This is an example of a **verse**:

ولا عيب فيهم غير أن سيوفهم - بهن فلول من قراع الكتائب

> There is no fault in them, excepting that their swords
> have suffered dents in clashes with battalions

> **Example 2.** A classical Arabic verse

Of the 180 rhetorical elements in Arabic poetry [Cachia 1998], we focus on the 20 most frequently used. In the following, we bring brief definitions, sometimes with simple examples. Detailed definitions, including detailed examples, are found in the Appendix (9.1):

1. *Simile* (تشبيه): comparing directly two objects; indicating that A is similar to B.

We have divided the following eight elements (nos. 2.1, 2.2, 2.3, 3.1, 3.2, 3.3, 4, 5) into four main groups(nos. 2–5), according to the identification method:

2. *Repetition1* (تكرار1), comprising three minor elements:

2.1. *Repetition* (تكرار): repeating the same word/s at least twice in the same verse while preserving the same meaning and the same grammatical structure and context (such as "we are here; we are here").

2.2 *Ploce* (ترديد): repeating the same word at least twice in the same verse; having the same meaning, but with a different grammatical structure/context (such as "we *are here*; they *are here* ").

2.3. *Perfect Paronomasia/Perfect Pun*; also Complete Paronomasia (جناس تامّ): Two words that are repeated in the same verse, each having a different meaning. The different meaning is a main difference between repetition and paronomasia.

3. *Repetition2* (تكرار 2), comprising three minor elements:

3.1. *Extraction/Etymology* (اشتقاق): Mentioning a personal name and another verb derived from this name — such as *Ṣāliḥ* (personal name) and *ṣaluḥa* ("becoming good").

3.2. *Semblance* (مشاكلة): A word is used in a certain, actual (often non-metaphoric) sense; it is then repeated in another, non-actual (often metaphoric) sense within the same verse (such as "we **breathe** the fresh air outside; while you remain in your room **breathing** the data from your books").

3.3. *Free paronomasia* (جناس مطلق): two words mentioned in the same verse, derived from the same root, or they look as if they have derived from the same root, and they have different meanings (such as *fiḍḍa* ("silver") and *faḍāʾ* ("space").

4. *Repetition3* (تكرار), echoing the rhyme at the beginning of the line (رد الاعجاز على الصدور): the final word in a verse is repeated (with the same meaning or with different meanings; with the same form or in a derivative form) anywhere in the first hemistich of the same verse or as the first word in the second hemistich such as (A....//....A; or ..../A....A).

5. *Repetition4*, key words (كلمات مفتاح): a root that is repeated in more than one verse in the same poem.

6. *Flowing* (اطّراد): praising a person by noting the names of his father and grandfather/s. The word *ibn* ("the son of") separates the names ("A is *the son of* B *the son of* C....").

7. *Afterthought/Retraction* (رجوع): mentioning one thought, then negating it in the same verse ("the abodes have not been changed through ages. Nay, I see that they have been totally demolished").

8. *Catchword Verbal* (تشابه الأطراف اللفظي): A verse opens with the same rhyming word which closed the previous verse ((1) A.... B; (2)B....C; (3)C....D).

9. *Distribution Characters/Alliteration* (توزيع)

9.1. *Distribution Characters/Alliteration 1* (توزيع 1): a single vowel or consonant is included in each word of a given verse.

9.2. *Distribution Characters/Alliteration 2* (توزيع 2): a certain vowel or consonant is repeated in all or most words of an entire poem.

10. *Hemming* (تصريع): the two hemistiches of a given verse end with the same rhyming letters.

11. *Unraveling* (توشيع): the verse ends with a general word in the dual form, followed by two single words that specify the two objects constituting this duality ("I am suffering from *two different situations: fear and hope*").

12. *Counterchange* (عكس, تبديل): two or more words appear in a certain order and are repeated in the same verse in reverse order (AB...BA).

13. *Repartee* (مراجعة): a conversation between different persons/objects in the same verse or in several verses in the same poem (He said.... I replied....).

14. *Rhyming – in general – homoeoteleuton* (سجع), dividing discourse into periods with similar-sounding last syllables. It is said to be:

14.1. *Congruent* (مُوازَن): the verse includes several phrases; their final words agree in measure and rhyme.

14.2. *Terminal* (مُطَرَّف): as in 14.1, but with the final words agreeing in measure only.

14.3. *Tucking* (تشطير): each of the two hemistiches in a verse is divided into two phrases that share the same rhyming letters. The rhyming letters in the phrases of the second hemistich differ from those in the first (A..A // B..B).

14.4. *Embroidery* (ترصيع): each word in the first hemistich shares the same rhyming letter (and sometimes the same measure) with its equivalent word in the second hemistich (ABCD//ABCD).

15. *Verbal Congruence* (مناسبة لفظيّة): the second hemistich includes a phrase that rhymes with an equivalent phrase in the first hemistich. The two phrases often open the two hemistiches.

16. *Paronomasia/Pun* (جناس), using or suggesting in a single verse words that differ in meaning but are phonetically and/or graphically alike or almost alike.

16.1. *Conjunct Paronomasia* (جناس مركب - مَرْفُوّ): the paronomasia consists of two parts: a single word (the first part) and two words (the second part) which, when combined, are similar phonetically and graphically to the first part of the paronomasia (*amradā* ("beardless") and *am radā* ("or a death?").

16.2. *Tipped Paronomasia* (جناس ناقص - مُطَرَّف): one of the matching words of the paronomasia is longer than the other with at least one initial consonant (*badā* "appear" and *abadā* "ever").

16.3. *Tailed Paronomasia* (جناس ناقص - مُذَيَّل): one of the matching words of the paronomasia is longer than the other with at least one ending consonant (*qanā* "lances" and *qabābil* "squadrons of horses").

16.4. *Consonantal/Distorted Paronomasia* (جناس محرف): the two matching words differ only in their vocalisation (and of course their meaning) (*ʿabra* "a tear" and *ʿibra* "a lesson").

16.5. *Substitutive or Variant Paronomasia* (جناس التصريف): the two matching words have a single differing consonant (*ilḥāḥ* "insistence" and *ilḥāf* "importuning").

16.6. *Metathetic/Reverse Paronomasia* (جناس مقلوب): all or some of the letters of one of the two matching words is in reverse order (*sāqin* "cup-bearer" and *qāsin* "cruel").

16.7. *Graphic Paronomasia* (جناس تصحيف): The two matching words, when written, are shaped alike, differing only in their dots which, in the Arabic script, differentiates some letters from others (*ḥabs* حبس "to restrain" and *jins* جنس "type").

17. *Negative Antithesis* (طباق سلب): a word is introduced and then negated with a negation particle ("A" and "not A" ("I know" and "I do not know")).

18. *Positive Antithesis* (طباق إيجاب): two different words that have contrary meanings ("new" and "old").

19. *The Satirist's Feint* (تأكيد الذّم بما يشبه المدح): Criticising an attribute in an individual, then exempting this by alluding that a positive attribute will be noted. The poet, however, disappoints expectations by referring to another negative feature rather than a positive ("he has no good attribute, *except of* being the worst person in universe").

20. *The Encomiast's Feint* (تأكيد المدح بما يشبه الذّم): praising an attribute in an individual, then exempting this, alluding that a negative attribute will follow. Here too, expectation is disappointed with another positive feature, instead of the expected negative ("they have no fault, except their swords which they broke on the heads of their enemies").

4. Tools and Methods

For the purpose of the study, a dataset of annotated poems was built. A set of 20 rhetorical elements was selected and the poems were annotated accordingly. The individual properties of each rhetorical element must be understood to create an automatic process that will recognize them. This identification requires different kinds of resources and techniques. Some features can be detected by applying one-to-one string-matching procedures, others by using Arabic morpho-syntactic analysis to tease out the "ingredients" of each word (e.g., root, stem), others still with sentimental analysis, which requires resources such as sentiment lexicons.

26

Working with only a small number of annotated poems, we were unable to use any classical machine-learning technique to identify these elements automatically, so our approach was rule-based.

27

A prototype of a rule-based system was built for automatic identification of these elements. The dataset, tools and methods are now described.

28

i. Poems Dataset

Over the years, we collected a repository of classical Arabic poetry[4] comprising 26,925 poems by 675 different poets. It includes pre-Islamic poetry, written before 622 CE; work by poets who

29

lived in the pre-Islamic and early Islamic periods, known as *mukhaḍrams*; that of poets who lived in the Umayyad era (661-750 CE); and of Abbasid poets who lived after 750 CE. For the purpose of this study, 75 poems, composed according to the classical Arabic meters called the Khalīlian meters, and comprising an average 27 verses each, were collected and manually analyzed to identify their rhetorical elements. The poems were chosen to cover two main periods: the Abbasid (between the 8th and 10th centuries CE) and pre-Abbasid eras (poets who lived between the 6th and 8th centuries CE). The poems of the first period are often shorter than those of the second, and therefore their number is greater: 48 poems in comparison with 29 from the first era (five pre-Islamic poems, seven poems which were composed by poets who lived in pre-Islamic times, but whose lifetimes extended into the early Islamic [*mukhaḍrams* poets], and 17 poems from the Umayyad era). The number of verses in the poems in the two main eras is, however, to some degree balanced. It comprises 887 verses from the Abbasid period and 752 from the pre-Abbassid (140 pre-Islamic verses, 232 *mukhaḍrams*, and 380 Umayyad).

From these poems, 20 different rhetorical elements were selected and manually annotated. (See Appendix 9.3 for the exact number of examples in each dataset.) `30`

ii. NLP Tools and Data Sources

This subsection describes the NLP tools and data sources, and the techniques applied in the framework. `31`

1. *AlKhalil2*, The AlKhalil [Boudchiche et al. 2017] was chosen because of its large database and because it is a morpho-syntactic analyzer of standard Arabic words out of context. The system analyses either partially or totally vowelized words. We used their morphological features, including prefix, stem and type of word, all of which are very useful in identifying most of the rhetorical elements. It is important to note, however, that AlKhalil2 was designed for modern Arabic whereas the classical Arabic poetry with which we worked differs from modern Arabic. As a first step, we report on evaluating it on our corpus and then discuss its implications.
2. *ArSenL Sentiment Lexicon*, The ArSenL [Badaro et al. 2014] was selected for its large sentiment lexicon (154,396 entries). They comprise the eight fields shown in our background section — positive sentiment score (3), negative sentiment score (4) and AWN lemma (8). When the AWN lemma includes NIL (i.e., not in language), we interpreted the morphological word in the first field according to the Buckwalter Arabic Morphological Analyzer [Buckwalter 2002b] to retrieve the Arabic word-form. Where the field AWN lemma included a valid offset (i.e., not NIL), we retrieved the Arabic word from AWN [Al-Khaṭīb 2003]. The final step was calculating the objective score of each word, which is the sum of the fields of positive and negative sentiment scores. It should be noted that most words scored neutral.
3. *Arabic WordNet*. We used the AWN source code to retrieve Arabic words from sentiment lexicons (described below).
4. *Buckwalter Arabic Morphological Analyzer* [Buckwalter 2002b] helped build a sentiment lexicon (tool 2), because entries in ArSenL [Badaro et al. 2014] contain fields that include NIL. This tool was used to convert Morphologic Arabic words to Arabic words (e.g., <imobiyriyAliy~ to إمبيريالي).
5. *Arabic Sentiment Lexicon*. We developed an Arabic sentiment lexicon, comprising 8,765 positive words and 9,813 negative words, in three phases. The first phase involved automatically retrieving Arabic words and their sentimental polarities from ArSenL [Badaro et al. 2014], using the Buckwalter Arabic Morphological Analyzer [Buckwalter 2002b] and AWN [Elkateb et al. 2006]. The second phase manually retrieved Arabic words and their polarities from another lexicon [ElSahar and El-Beltagy 2015], and the final phase was manually enriching the lexicon with Arabic words and their polarities from the internet.[5]
6. *Arabic Sentiment Analyser*. We developed a word polarity-based Arabic Sentiment Analyzer using the lexicon described above (see Figure 7).
   - Step 1 was building a lexicon of sentiment words. It comprised two text files — the first file consisting of positive words, the second of negative. In addition, we manually created a text file of negative words.
   - Step 2 was building an Arabic Sentiment Analyser based on polarity. Text verses were inputted, each verse devoted to one word in all its possible forms. (All words in each verse appear without diacritics because the lexicon includes words without diacritics.) The analyzer checks each word in the verses. Should it appear in one of the lexicons, its polarity value is set according to that lexicon. If, however, there is a negation particle before the word (in Arabic, they are mainly بلى, ليس, لكن,لم، لا، لن), its polarity is assigned the opposite value of the lexicon. The analyzer then calculates the final sentiment of the input: if the number of positive words is greater than that of negative words, the analyzer returns a value of 1; if negative words comprise the greater number, it returns -1; where they are equal, it returns 0 (i.e., neutral). This serves to identify the Satirist's and Encomiast's Feints (see Figure 6).
7. *Arabic Lexicon for Antithesis*. We developed an Arabic lexicon for words with contradictory meanings. It comprises 453 pairs of opposites, compiled in two stages. First was retrieving English words and their opposites from the internet ([Englisch-hilfen.de n.d.] [Perfectyourenglish.com 2019]), and using the AWN Browser 2.0.1 to get their Arabic opposites and synonyms. Second was manually enriching these opposites with Arabic words and their synonyms [Al-Jāḥiẓ 2019].
8. *Graphical User Interface (GUI)* (see Figures 2 and 4).

iii. Rhetoric Elements Identification Framework and Prototype

We created a user-friendly system that integrates the NLP tools described above with a rule-based elements identification mechanism for automatic identification of rhetorical elements. The GUI and the algorithms are from NetBeans IDE 8.0.1 using Java platform JDK 1.8. The source codes of the first six tools were also implemented in Java. Figure 1 presents the abstract process of rhetorical element identification on a complete poem, line by line, while Figures 1 & 5 show the process of identifying a rhetorical element. Figure 1 shows the overall abstract process of line by line analysis while Figure 5 shows an example of identifying one specific rhetorical element – "Simile" (as there are different implementations for different rules). `32`
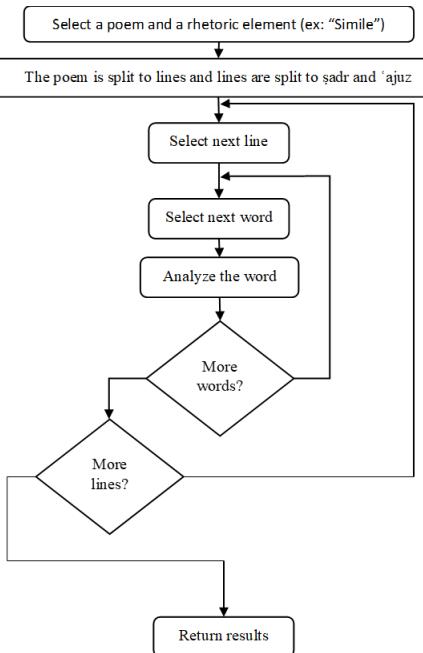

Figure 1. Abstract framework for rhetorical element identification. The process starts with selecting a poem and a rhetorical element. The system analyses the poem line by line to identify the rhetorical element.

The *Graphic User Interface* (GUI) for identifying rhetorical elements is shown in Figure 2. The process is as follows (see Figure 3):

1. Activate the system (1), and select a poem (2).
2. Select a rhetorical element from the 20 in the GUI.
3. Run the tools to identify the selected element for each verse. If more than one element is chosen, repeat the process for each
4. The color-coded results are visualized (3).
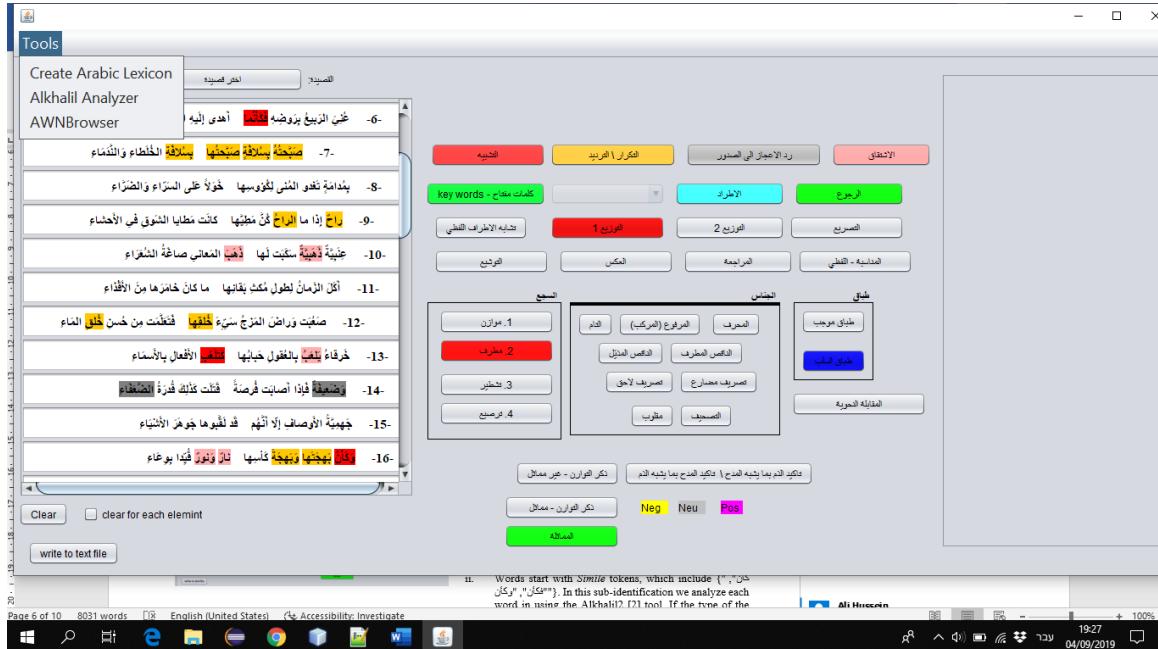
Figure 2. The main Graphic User Interface (GUI). The selected poem is on the left. Each button on the right is responsible for identifying a rhetorical element, color-matched to the poem's text.
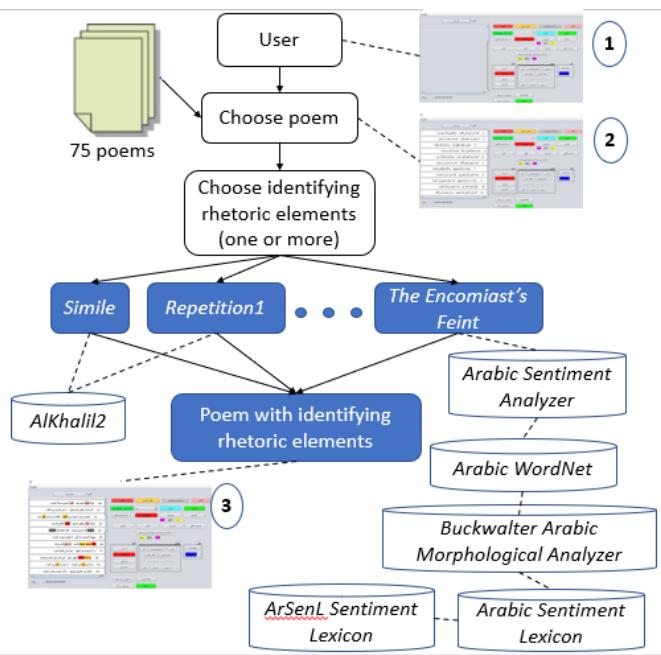
Figure 3. Illustration of the automatic annotation framework — from when the user initiates the process (1) to the desired result (3).

We also created a GUI for the Arabic Sentiment Lexicon Tool, enabling building of the lexicon and interpreting the morphology of words (Figure 4).
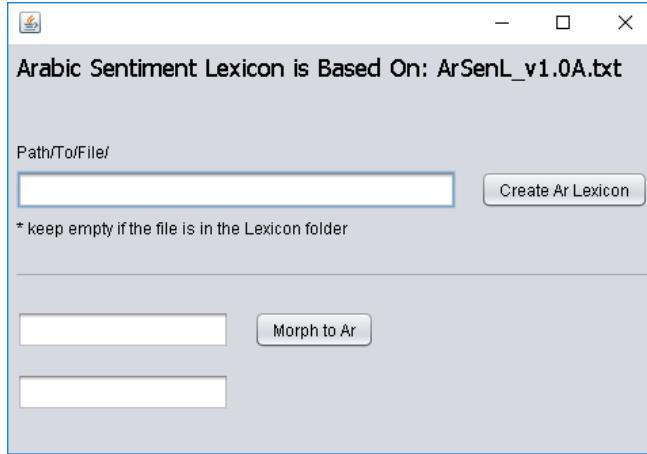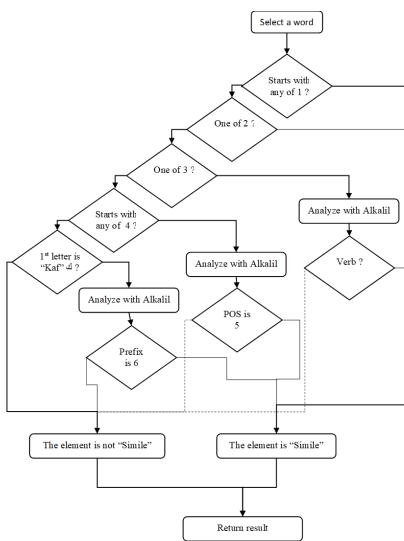
Figure 4. GUI for building the lexicon user file and interpreting the morphology of words.

Figure 5. Rhetorical element identification process example: Simile. The process is repeated for each word in a line, following a set of conditions that refer to the start of the word or whether it is one of a set of predefined words. If needed AlKhalil is consulted for morphological analysis and the results are further checked.
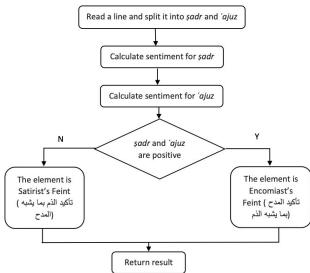
Figure 6. Rhetorical element identifying process example, Satirist's Feint and Encomiast's Feint: The process starts with checking the sentiments of the *ṣadr* and *ʿajuz*, word by word, and then combining the results.
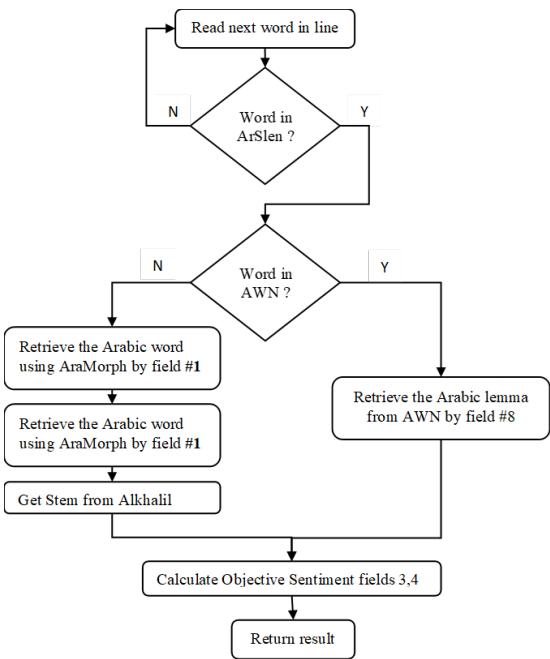
Figure 7. Identifying a words' sentiment.

Another essential component enabled us to identify the emotions expressed by words (Figure 7), and, when needed, to calculate the overall emotion expressed by a complete line (Figure 6).

iv. Identifying Rhetorical Elements (Rules)

The individual properties of each rhetorical element must be understood to create an automatic process capable of recognizing them. This identification requires different kinds of resources and techniques. Some features can be detected by applying one-to-one string-matching procedures, others by using Arabic morpho-syntactic analysis to tease out the "ingredients" of each word (e.g., root, stem), others still with sentimental analysis, which requires resources such as sentiment lexicons. Next, we describe the rules we applied to identifying each rhetorical element, how we applied them, and the tools we used.

1. *Simile* (تشبيه). This can appear with and without word tokens. The latter is not included in our system. This rhetorical element is identified in two steps. The first identifies it by full word tokens in one of four groups (see also Figure 5):

1.1. Words starting with the tokens مثل, فمثل, ومثل, كالـ for which the system searches.

1.2. Words starting with the simile tokens كأن, فكأن, وكأن. Each word is analyzed using the AlKhalil2 [Boudchiche et al. 2017] tool. If the type of the word is a حرف مشبّه بالفعل (particle that resembles the verb; a group of particles including إن – أن / كأن / ليت / لكن – لعل) and the word is one of the three tokens mentioned above, it is considered a simile.

1.3. All word tokens كما, كأنما, فكما, وكما, وكأنما, فكأنما. The system searches for whole tokens. For each, vocalization (تشكيل) is removed, and it is then checked as to whether it is equal to one of the tokens. Where a match is found, it is a simile.

1.4. Verb tokens ظن, خال, حسب, شبه are analyzed using the AlKhalil2 [Boudchiche et al. 2017] tool to retrieve type and root. If its type is a verb and its root is equal to one of the tokens, it is a simile.

The second step is identifying similes by the letter k- ـكَ. Each word is analyzed using AlKhalil2 [Boudchiche et al. 2017] to check whether it is a prefix. If the prefix is equal to :ـكَ حرف التشبيه, it is a Simile.

2. *Repetition 1-4* (تكرار). These rhetorical elements are identified by two tests. The first is when two identical words appear in the same verse. This is Repetition. The second is finding the same stem in more than one word, with different predefined prefixes. We exclude words of the following types: { حرف جر – preposition, حرف نداء – interjections, حرف – disjunctive, and حرف عطف – copulative particle, and particles that indicate حرف نصب – accusative, حرف نفي – negation, حرف جزاء – compensation, حرف عطف – disjunctive, and حرف جزم – apocopative, أداة شرط - conditional particle}. Such particles are frequently repeated in poetry and are not, therefore, considered part of the text's rhetorical texture. To exclude these words, we analyze every word in the verse using AlKhalil2 [Boudchiche et al. 2017] to determine its type. Where the type matches one of those mentioned above, the word is excluded. These tests also eliminate words used in the "flowing" "rhetorical element no. 6" identifications. To identify Repetition1-4-type (see Appendix 1) rhetorical elements, contextual association is also needed — for example, Semblance (مشاكله). At this point, however, it is sufficient to recognise repetition of the same word without reference to context.

3. The *Flowing* (اطراد) is when the word بن appears at least twice in the same verse.

4. *Afterthought/Retraction* (رجوع) is when the first word in the second part of the verse matches one of the predefined negation words — وما - "and it is not," بلى - "yes, it is!," وغير – "but it is".

5. *Catchword Verbal* (تشابه الاطراف اللفظي) is when the last word in a verse reappears as the first word of the following verse.

6. *Distribution Characters/Alliteration* (توزيع). Type 1 is when the same consonant appears in all or almost of all words of a given verse. Type 2 is when the same consonant appears in all or most of the verses in the poem. The vowels و ي, l are ignored.

7. *Hemming* (تصريع) compares two-word suffixes — the last word in the first and second parts of the verse.

8. *Unraveling* (توشيع) is when the verse ends with a phrase containing a noun in dual form (مثنى), followed by two singular nouns connected with the particles *wa* or *fa* ("and"). Each word in the second part of the verse is analyzed with AlKhalil2 [Abdul-Mageed, Diab, and Kübler 2013] to check whether it is dual form.

9. *Counterchange* (عكس وتبديل) is when two sentences are composed of the same words, but in reverse order.

10. *Repartee* (مراجعة) is when two consecutive verses contain a reported conversation. The system finds this by searching for the word "said" and its derivatives.

11. *Rhyming – in general – homoeoteleuton* (سجع) is when the concluding syllables (the end of two parts) in two sentences within the same verse end with the same consonants. Where they sound similar, it is a *saj'*.

12. *Verbal Congruence* (مناسبة لفظيّة) is when the first and second parts of the poem contain individual words in the same locations that resemble one another, either in terms of their suffix (their final letter) or in their metrical pattern, or in both.

13. *Paronomasia* (جناس) comprises seven types that are identified in two phases. First, the prefixes and suffixes of each word are removed, and then the letter, vocalization and type of word-pair are examined.

14. *Negative Antithesis* (طباق سلب) is when two words (the same words or words that are derived from the same root) are repeated, and a negation particle precedes one of them.

15. *Positive Antithesis* (طباق إيجاب) is identified using our *Arabic Lexicon of Opposites*, which lists words and their opposites. It searches for words and their opposites within the same verse.

The following two rhetorical elements are based on sentiment analysis. To identify them, we used an *Arabic Sentiment Analyser*.

16. *The Satirist's Feint* (تأكيد المدح بما يشبه الذّم) searches each verse for predefined terms — أن لا, أن يوس , أن ريغ ان يوس , أن ريغ , ان ريغ ,. Where found, the sentiment values of the words that precede and follow the term are calculated separately. When both values are positive, it is a Satirist's Feint.

17. *The Encomiast's Feint* (تأكيد الذّم بما يشبه المدح) is calculated as above. When both sentiment values are negative, it is an Encomiast's Feint.

5. Experiments and Results

This section first describes the study that evaluated the suitability of AlKhalil2 for our task, and then the datasets, the process of defining the rules and testing the framework. Figure 8 illustrates the evaluation process. [36]

i. Dataset

As noted, we used a dataset comprising 75 classical era Arabic poems, composed by 35 poets (see Appendix 2 for the names and numbers of poems). [37]

We divided these poems into two datasets: a training set (77%) and a test set (23%) (Appendix 2 shows details of poems and rhetorical elements in each). All were then manually annotated by an Arabic literature researcher. Each poem consists of several verses that include different rhetorical elements (see Appendix 4). It is important to note that we used only the training set for developing the rules for identifying the different rhetorical elements, while the test set was used to test the system. The experimental results described later refer to the automatic analysis of examples that were set aside and not used for training (development of rules). [38]

The datasets are small because of the intensive manual effort required in annotation. There is, thus, only a modest number of manually tagged poems, each containing a modest number of rhetorical elements. Rules for identifying elements, which were not found in any of the poems (see Appendix 9.4), were defined by creating a second dataset of verses from poems in which they are featured ([Habash and Rambow 2006], [Ahmed and Trausan-Matu 2017], [Diehl et al. 2009]). [39]
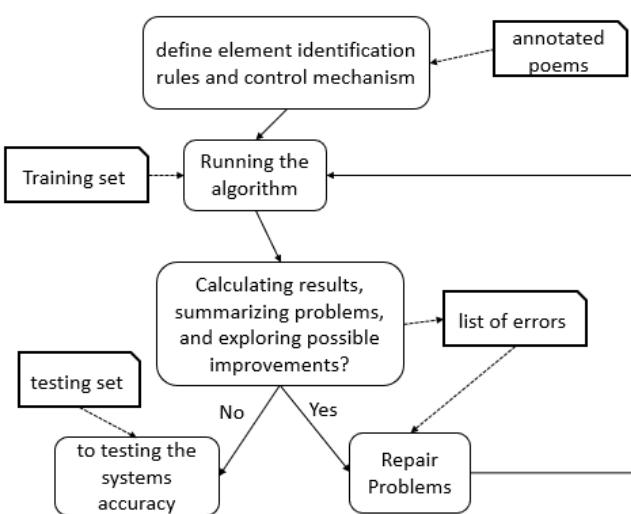

Figure 8. The evaluation procedure: defining and testing rules to build algorithms with maximum success.

ii. Procedure

The procedure (Figure 8) was: (1) defining identification rules by studying the training set; (2) testing the system on the training set to debug, correct, improve and refine the rules, repeating the process until results could not be improved further; (3) finally, we applied the resulting rule-based system to the test set, that had been set aside and not used for training, and measured the results with precision, recall and F1 measures [Sasaki 2007].

iii. Evaluating AlKhalil2

Given the potential language differences, we first evaluated the suitability of AlKhalil2 as an NLP tool.

We thus selected eight poems from our 75 poems by different poets, regions and times, used AlKhalil2 to analyze them, and checked the results manually. Overall, the results were good, but far from perfect (Table 1 presents the results).

Overall, AlKhalil2 correctly analyzed 74.2% of the words. It did not analyze 25.8% or analyzed them wrongly. The majority of the unanalyzed/wrongly analyzed words are either proper names (places, plants, tribes or given names) unknown to the system, or words which are incompletely or improperly vocalized. By improperly we mean one of four possibilities: (1) a consonant of the given word has no diacritics (i.e,, is not vocalized); (2) the "Sun" letters (*t, th, d, dh, r, z, s, sh, ṣ, ḍ, ṭ, ẓ, n*), that follow the definite article *al-*, are not vocalised with the *shedda* mark (indicating that a consonant is duplicated); (3) the *tanwīn al-fatḥ* is typed over the vowel *ā* instead of over the preceding consonant (such as كِتَابًا instead of كِتَابًا meaning "a book" in the accusative form); both ways of putting the *tanwīn* are possible in Arabic, however, in AlKhalil2 the second way is preferred); (4) the rhyming letter, in the accusative form, is often suffixed to the *ā* vowel, which is a prolongation of the *fatḥa* (such as *akala* is written a *akalā*; the verb means "to eat").

Poem Details	Number of Words	Correct	Incorrect
Imruʾ al-Qays poem no. 8	194	131 (67.5%)	63 (32.4%)
Al-Aʿshā Maymūn 78	224	153 (68.3%)	71 (31.7%)
Abū Dulāma (1)	116	84 (72.4%)	32 (27.6%)
Al-Najāshī l-Ḥārithī 41	116	85 (73.2%)	31 (26.7%)
Jarīr 11	195	173 (88.7%)	22 (11.3%)
Ibn Qays al-Ruqayyāt 50	207	135 (65.3%)	72 (34.7%)
Dīk al-Jinn al-Ḥimṣī 38	86	72 (83.7%)	14 (16.3%)
Al-Namir b. Tawlab 34	195	156 (79.9%)	39 (20%)
Total	**1,333**	**989 (74.2%)**	**344 (25.8%)**

Table 1. AlKhalil2 performance on classical Arabic poetry

AlKhalil2 wrongly identifies this *ā* as the dual pronoun *ā*. In addition, in poetry, because of metric considerations, sometimes the masculine *y* and feminine *t* prefixes, from the imperfect verb in the two forms *tafaʿʿala* and *tafāʿala* (known as forms 5 and 6), are omitted (i.e., *yatafaʿʿal* and *tatafaʿʿal* can become *tafaʿʿal*; while *yatafāʿl* and *tatafāʿal* becomes *tafāʿal*). In these cases, AlKhalil fails to analyse the verb. Most can be addressed by a manual revocalisation or retyping. In any case, while the percentages 25.8 and 8.4 make AlKhalil2 less than perfect, its results were currently good enough for us (especially since it was the best tool we had).

iv. Framework Evaluation Results

The overall results were quite good: overall precision of 0.918, recall of 0. 902 and F measure of 0.906. They are presented in Table 2. The results differ greatly, with extremely successful identification of most of the elements, but relatively lower success levels in identifying Negative Antithesis, Positive Antithesis and the Satirist's and Encomiast's Feints. These differences are further discussed later.

We measured the success rate in identifying each rhetorical element separately, other than for Repetition1, Repetition2, the Satirist's Feint and the Encomiast's Feint (the results are presented in Table 2). Elements 2.1, 2.2, and 2.3 (Repetition1) were treated as a single element since their classification is based on the same general rule with only minor differences — in 2.1, the word is repeated in the same context; in 2.2, two words with the same meaning are repeated in two different contexts; 2.3 resembles 2.2 but the meaning of the repeated words differs.

	Precision	Recall	F1
Simile (التشبيه)	0.916	0.891	0.904
Repetition1 (التكرار)	0.962	0.838	0.896
Repetition2 (الاشتقاق)	0.902	0.925	0.913
Repetition3 (رد الاعجاز على الصدور)	0.941	0.941	0.941
Repetition4 – Key Words	1	1	1
Flowing (الاطراد)	1	1	1
Afterthought/ Retraction (الرجوع)	1	1	1
Catchword Verbal (تشابه الاطراف لفظي)	1	1	1
Distribution Characters/Alliteration 1 (التوزيع)	1	1	1
Distribution Characters/Alliteration 2 (توزيع 2)	1	1	1
Hemming (تصريع)	1	1	1
Unraveling (توشيع)	1	1	1
Counterchange (عكس وتبديل)	1	0.8	0.889
Repartee (مراجعة)	1	1	1
Rhyming – in general – homoeoteleuton (سجع)	1	0.875	0.933
Verbal Congruence (مناسبة لفظيّة)	0.778	0.875	0.824
Paronomasia (جناس)	0.857	0.75	0.8
Negative Antithesis (طباق سلب)	0.636	1	0.777
Positive Antithesis (طباق إيجاب)	0.551	0.457	0.5
Satirist's, Encomiast's Feint تأكيد الذّم بما يشبه المدح تأكيد المدح بما يشبه الذّم	0.818	0.692	0.75
Average	0.918	0.9o2	0.906

Table 2. Results of the rhetorical element identification system. T= true; F= false; P = positive; N = negative.

Rhetorical elements 3.1, 3.2 and 3.3 (Repetition2) are also based on the same main rule with minor differences. Rhetorical elements 19 and 20 (Satirist's Feint and the Encomiast's Feint) are treated as one, since they follow the same rules other than their difference in polarity. Table 3 shows the true positives, false positives and false negatives derived. There were relatively high numbers of false positives and false negatives for only two of the rhetorical elements — Positive Antithesis (طباق ايجاب) and Paronomasia (جناس).

	T-P	F-P	F-N
Simile (التشبيه)	33	3	4
Repetition1 (التكرار)	26	1	5
Repetition2 (الاشتقاق)	37	4	3
Repetition3 (رد الاعجاز على الصدور)	16	1	1
Repetition4 – Key Words	45	0	0
Flowing (الاطراد)	3	0	0
Afterthought/ Retraction (الرجوع)	5	0	0
Catchword Verbal (تشابه الاطراف لفظي)	8	0	0
Distribution Characters/Alliteration 1 (التوزيع)	5	0	0
Distribution Characters/Alliteration 2 (توزيع 2)	56	0	0
Hemming (تصريع)	5	0	0
Unraveling (توشيع)	14	0	0
Counterchange (عكس وتبديل)	4	0	1
Repartee (مراجعة)	7	0	0
Rhyming – in general – homoeoteleuton (سجع)	21	0	3
Verbal Congruence (مناسبة لفظيّة)	7	2	1
Paronomasia (جناس)	18	3	6
Negative Antithesis (طباق سلب)	7	4	0
Positive Antithesis (طباق ايجاب)	13	16	19
Satirist's, Encomiast's Feint تأكيد الذّم بما يشبه المدح تأكيد المدح بما يشبه الذّم	9	4	2

Table 3. Success rate. T= true; F= false; P = positive; N = negative

6. Discussion

Our results demonstrate that the system performs extremely well for most rhetorical elements — Repetition4, Flowing, Afterthought/Retraction, Catchword Verbal, Distribution Characters/Alliteration, Hemming, Unraveling and Repartee (Table 2). This can be explained by the fact that identifying these elements requires checking the words and their occurrences without analysis or consideration of context. This means that identification of these rhetorical element is not conditioned by recognition of the context in which they appear, so the system does not need to understand the meaning of the verse to decide whether these elements are present. For example, opening a verse with the same word in which the previous one ended (the rhetorical element known as the Catchword Verbal) can be noted without understanding the content of the verses in which they appear.

Other rhetorical elements proved more challenging with errors in their identification. We conducted error analysis for each, and suggest possible solutions to improve results. This remains for future study.

While the framework's identification of Simile and Repetition2 is high (precision = 0.916; recall = 0.925; F1= 0.913), we assume that the second test in Simile identification (that is, by the prefix *ka-* which is the particle of simile meaning "like". The word after it is in the genitive) can be upgraded by considering context. Using the AlKhalil2, we encountered situations where a word could be segmented into multiple forms — one of the forms with the simile particle *ka-*, and the other with the *ka-* as part of the word. The word كواعب, for example, can be analysed in multiple forms, one of them with the simile *ka-: ka-waʿib*; i.e., "similar to a person who takes everything"), but considering the context of the verse (which describes women), the word should not be understood as a simile; the *k* is part of the word *kawāʿib* (pl. of *kāʿib*) which means "girls with swelling breasts". Moreover, identification of Counterchange, Rhyming and Paronomasia can be improved when we have sufficient poems containing them, allowing definition of appropriate rules. The current set is drawn from the scarce appearances of these rhetorical elements in the training set.

Identification of Repetition1 and Repetition3 is very good but can be further improved by considering context. Because the false negative rate for Repetition1 is high, we must add an option for removing them. Negative Antithesis identification, too, can be bettered by adding negation words, with further study needed to resolve situations where a word appears between the negation word and that which is repeated — for example, in poem *Al-Najāshī* 62:

وَمَا زَالَ مِنْ هَمْدَانَ خَيْلٌ تَدُوسُهُمْ *** سِمَانٌ وَأُخْرَى غَيْرُ جِدَّ سِمَانِ

The horses of the Hamdān tribe continued to trample upon them. These include both **fat** and **not-fatted** horses.

Example 3. A passage from *Al-Najāshī* 62

Identifying the Satirist's and Encomiast's Feints was less successful for several reasons: (1) a word may have positive and negative polarity, which requires contextual analysis (that is, guessing the verse's meaning according to its other content). (2) The lexicons include words retrieved from non-formal sources, such as Twitter, requiring creation of a richer lexicon based on formal language. (3) The poet sometimes boasts of negative qualities or facts, as if they are positive (e.g., "His sword is dripping blood in war" means "he is a warrior"), which requires consideration of context. And (4) ambiguous words can be misclassified, again requiring attention to context. Lowest of all in performance is identification of Positive Antithesis, largely because the *Arabic Lexicon for Antithesis* which we built is insufficient. Since false negatives rate high in this poetry, a new, richer *Arabic Lexicon for Antithesis*, based on formal language, must be created.

The diacritical text in which classical poems are usually written increases the accuracy of text analysis, reducing, as it does, the number of possible segmentations for each analyzed word. When this same text is inputted into the framework without diacritics, results are poorer. The AlKhalil2 tool returns all possible diacritical options for the word.

As in any study, ours has limitations, some of them indicated above. The first is its relatively small number of poems — 75 of the total 26,925 classical Arabic poems in the canon. The second is that some of these 75 are written without diacritics. And third is that the AlKhalil2 misses small parts — for example, the word كتيس in the verse:

مِكَرٍّ مِفَرٍّ مُقْبِلٍ مُدْبِرٍ مَعًا *** كتيس ظِباءِ الْحُلَّبِ الْغَذَوَانِ

Ready to charge, ready to fell, advancing, retreating equally well; [its speed] like [that of] a lively buck-gazelle that feeds upon the *ḥullab* trees, rock brought down from on high by [a raging] torrent.

Example 4. A classical Arabic verse, read by AlKhalil2

This leads to a fourth limitation and is directly related to AlKhalil2 performance on classical Arabic poetry that is far from its performance on modern Arabic. Adapting it to classical Arabic poetry is an idea for future research.

55

Finally, the lack of a large annotated corpus forced us to apply a rule-based approach that is limited by the examples annotated to date, and considered by the developers. The rule-based approach is a well-known method that was used in early days of expert systems. It has numerous limitations, as it depends heavily on manual encoding of reasoning rules. It also applies only to instances for which rules were explicitly written, and thus needs ongoing maintenance [Hayes-Roth 1985]. This is a major limitation, especially given the small dataset we used. The process we followed exemplifies both the benefits of the approach as well as its limitations (also shown in the analysis above). Still, as noted, we see this as an interim solution, despite our very encouraging results. There remains great diversity between different rhetorical elements. We continue to strive for better results that ensure high quality results, even even sacrificing precision (accepting false positives) in order to increase recall (reducing, even eliminating false negatives). We thus suggest its use only until a sufficiently large dataset of annotated poems is available for training machine-learning classifiers for the task.

56

7. Conclusion and Future Work

Identification of rhetorical elements in classical Arabic poetry is a time-consuming process. It requires many resources, and high-level Arabic literature experts. A new, automatic, rule-based mechanism for their identification is proposed in this article, and its use with 20 elements is demonstrated as a case study. For most, the framework's performance is very high.

57

Further work is needed in several areas. One is addressing the study's limitations, presented in the Discussion, and exploring the suggested solutions, such as including a significantly larger number of manually analyzed poems in the system. Another is improving the performance of the current framework by adding rules to identify missed elements, ignore misidentification and generally upgrade results. A third direction is building Arabic Sentiment Lexicons based on formal words, which will manually enrich the Negation Word Lexicon and include context-based analysis.

58

This study selected 20 rhetorical elements for identification. The automatic identification of some of them (such as the Positive Antithesis, the Satirist's Faint and the Encomiast's Feint) can be improved by developing the existing program. There are others still to add — among them, Metaphor (the most challenging rhetorical figure), Periphrasis, Exaggeration and types of Paronomasia that we have not addressed — which require resources such as antonym lexicons and context-based analysis. One major impediment is that identifying these elements depends mainly on semantics: the system must be capable of recognising the semantic meaning of a word/expression/phrase in order to identify some of these rhetorical elements. Automatic identification of elements not addressed in the present research and the improving automatic identification of those that are, are left to future research.

59

The present system may also be used for identifying and studying rhetoric in the modern Arabic poetry. For this, poems from the 19th-21st centuries should be added to Hussein and Kuflik's existing system. Since most rhetorical elements in Arabic culture have equivalents in other literatures (the simile for example is also used in English, German, Chinese and more), the system can be adapted, modified and improved to identify and analyze rhetoric in non-Arabic literatures. Global use will enable in-depth comparisons of rhetorical fabrics across literatures and cultures.

60

Finally, we note that the proposed approach is generic, and the same framework may be applied to any kind of literature. The issue is not the framework's generality but the expertise, time and effort required for defining rules. We thus reiterate that the suggested approach may be used as an interim solution when there are insufficient examples of rhetoric elements to train a classical machine-learning algorithm.

61

Appendices

Appendix 1: Definitions of the 20 Rhetorical Elements (based mainly on [Cachia 1998])

The definitions, most examples and verse translations (sometimes with modifications) are taken mainly from Cachia [Cachia 1998]. It is, to the best of our knowledge, the only existing bi-lingual dictionary for Arabic rhetoric. As the author indicates in the introduction to the book's English-language edition, each rhetorical element has been given an English name although in some cases the Arabic element cannot be translated (p. 3). Where there is an identical or very similar rhetorical element in English, the English term is adopted. Where there is no or only partial equivalence, the term is translated by Cachia to reflect the meaning of the Arabic. The explanations of the terms clarify what it is intended by them.

62

1. *Simile* (تشبيه)
The indication, by the use of the word "like" ("similar") or some such word (مثل، كأن، كأنما، كما، كـ، شبه، حسب، خال), whether an object or condition explicitly or implicitly shares an attribute with another. The former object or condition is the tenor or *primum compartionis*, the latter is the vehicle or *secundum comparationis*. This is an example from the Syrian poet Abū Tammām (d. 846 C.E.):

> خَلَطَ الشجاعةَ بالحَيَاء فأصْبَحا ... كالحُسن شِيبَ لِمَعْرَم بدلال
> He mingled **bravery** (*shajāʿa*) with **reticence** (*bi-l-ḥayāʾi*) so that it seemed like (*ka-*) **beauty** (*l-ḥusni*) joined to **coquetry** (*dalāli*) in a lover's eyes

2. *Repetition1* (تكرار 1)
This comprises two minor rhetorical elements:

    2.1. *Repetition* (تكرار) is use of the same word or words, more than once, in the same form, and with the same meaning and context for the sake of emphasis in description, praise or another purpose. This example is from ʿAbd al-Ghanī al-Nābulsī (d. 1731 C.E.):

> والجِسْمُ والجِسْمُ قدْ أوْدَى السّقامُ به ... والجَفْنُ والجفنُ طولُ اللّيل ما هَجَعا
> **My body** (*wa-l-jismu*) — **My body** (*wa-l-jismu*) is reduced by languor, **My eyes** (*wa-l-jafnu*) —**my eyes** (*wa-l-jafnu*) have known no sleep all night

    2.2. *Ploce* (ترديد) uses the same word more than once, each time with a different application or context (sometimes it is a different grammatical context; that is, each of the two repeated words is attributed to a different subject/object/pronoun etc.). This, from Sayf al-Dawla, Emir of Aleppo (d. 967 CE), describes a slave-girl whom he was sending to a fortress to protect her from the jealousy of his other favorites:

> ... رُبَّ هَجْر يكونُ مِنْ خَوْف هَجْر وفِراقٍ يكونُ خَوْفَ فِرَاق
> **A break** (*hajr*) may be due to the fear of **a break** (*hajr*), And **parting** (*firāq*) be caused by the fear of **parting** (*firāq*).

    2.3. *Perfect Paronomasia/Pun* (جناس تامّ; also *Complete Paronomasia*) is the use in a single context of two words pronounced and written exactly alike but carrying different meanings. This example is from the poet Abū Tammām:

> مَا مَاتَ مِنْ كَرم الزّمَان فإنّه ... يَحْيَا لَدَى يَحْيَى بْن عَبْد الله
> All generosity that have perished is still **living** (*yaḥyā*) in the court of **Yaḥyā** the son of Abdallah

3. *Repetition2* (تكرار2)
This comprises three minor rhetorical elements:

    3.1. *Extraction* (اشتقاق; also *Etymology*) draws from a personal name an idea that serves the poetic motif, satirical, panegyric, erotic or other. This example, from Abū l-Ḥasan ʿAlī ibn Muḥammmad al-Anṭākī (d. 988 C.E.), is to a patron called Ṣāliḥ:

> يا صالحَ الخيرات ما صَلَحا ... إلا لك التأييد والأمْرُ
> O **Ṣāliḥ**, [the name of the patron], the patron of all bounties, the two virtues of ordering and being obedient **befit** (*ṣalaḥa*) only your personality

    3.2. *Semblance* (مشاكلة) replaces the denotative of an object or action with another word for reasons of contextual association. An example from Abū Tammām:

> والدّهْرُ ألأَمُ مَنْ شَرِقْتَ يَلومِه ... إلا إذا أشَرقْتَهُ بكَريم
> Fate is the vilest of all that **chokes you** (*shariqta*) with vileness,

Unless **you choke it** (*ashraqtahu*) with [the succour of] a generous man

3.3. *Free Paronomasia* (جناس مطلق) uses or suggests in one context words that differ in meaning but are phonetically and/or graphically alike or nearly alike — that is, matching words differ in their letters and vocalisation but have at least two radical consonants in common, sometimes falsely making them appear derivatives from the same root. (Sometimes they do derive from the same root). For example:

ذَهَبٌ حَيْثُما ذَهَبْنا ودُرٌّ ... حَيْثُ دُرْنا وفِضَّةٌ فى الفَضاء

Gold (*dhahab*) wherever **we go** (*dhahabnā*), **pearls** (*durr*) Wherever **we roam** (*durnā*), and **silver** (*fiḍḍa*) even **in space** (*fī l-faḍā'*).

4. *Repetition3* (تكرار 3)

This echoes the rhyme at the beginning of the line (رد الاعجاز على الصدور). It takes two words — which are identical in pronunciation and meaning, or almost alike in pronunciation but not in meaning, or derived (either genuinely or apparently) from the same root — placing one near the beginning of the discourse and the other at its end. This example is from Ibn al Fāriḍ (d. 1235 C.E.):

يا ساكنى البطْحاء هلْ من زورةٍ ... أَحْيا بها يا ساكنى البطْحاء

Desert dwellers (*yā sākinī l-baṭḥā'i*), will you not allow one visit That I may live by it, **O desert dwellers** (*yā sākinī l-baṭḥā'i*)?

5. *Repetition4 — Key Words* (كلمات مفتاح)

We suggest a new rhetorical element, in which a root is repeated in different verses of the same poem. While this includes Repetition1 and 2 (repetition, ploce, extraction/etymology, semblance and paronomasia), we identify it in all verses of the poem.

○ ### 6. *Flowing* (اطراد)

This features in a single verse the given personal name of the individual praised, as well as the name of his father and grandfather and that of the tribe, all in the appropriate order. This example is from Ibn Durayd al-Azdī (d. 933 C.E.):

عِيادُ بْنُ عَمْروبْنِ الحَلِيس بن جابر بْن زَيْدِ بْن مَنْظُورِ بْن زَيْدِ بْن وارثِ

Iyād, the son of (*bnu*) `Amr, the son of (*bni*) al-Ḥalīs, **the son of** (*bni*) Jābir, **the son of** (*bni*) Zayd, **the son of** (*bni*) Manẓūr, **the son of** (*bni*) Zayd, **the son of** (*bni*) Wārith (the *u* and the *i* at the end of the word indicate two different grammatical forms — nominative for the first, and genitive for the second).

7. *Afterthought* (رجوع; also translated as *Retraction*)

This reverses an earlier statement. The first word in the second part of the verse usually matches one of the predefined words {بل, بلى, ليس, لكن, نعم. لم, لا, لن}. This example is from the pre-Islamic poet Zuhayr ibn Abī Sulmā (d. 609 C.E.):

قفْ بالديار التى لم يَعْفُها القِدَمُ ... بَلى وغَيّرها الأرواحُ والدِّيمُ

Halt by a deserted abode that age **has not obliterated** (*lam ya'fihā l-qidamu*)–
Nay (*balā*), that winds and rain have altered!

8. *Catchword Verbal* (تشابه الأطراف اللفظيّ)

This is of two types. We consider one type only — the verbal — which corresponds to anadiplosis, gradatio, reduplication. It starts each of a succession of verses with the preceding rhyme word, for example, from the poetess Laylā l-Akhyaliyya (d. 700 C.E.):

إذا نَزَل الحجّاجُ أرْضًا مريضَةً ... تَتَبّعَ أقْصى دائها فشَفاها
شَفاها من الدّاء العُضالِ الذي بها ... غُلامٌ إذا هَزّ القَناةَ سقاها
سقاها قَرَواْها بشرْبِ سِجالَها ... دماءُ رجالٍ يَحْلِبونَ ضَراها

Whenever al-Ḥajjāj visits a sick land,
He treats its illness and **cures it** (*shafāhā*);
It is cured (*shafāhā*) of its chronic disease
By a warrior who always shakes his spear, and **quenches the land's thirst** (*saqāhā*);
He quenches the land's thirst (*saqāhā*) by drinking buckets of the blood of his enemies.

9. *Distribution Characters* (توزيع; also *Alliteration*)

9.1 *Distribution Characters/Alliteration 1* (توزيع 1)

This distributes to the poet or speaker the same consonant or vowel in every word of a verse. The programme checks for such a letter. This verse is by Ṣafiyy al-Dīn al-Ḥillī (d. 1349 C.E.):

محمدٌ المُصطَفى المُختارُ مَنْ خُتِمَتْ ... بمجدِهِ مُرْسَلوالرَّحْمَن للأُمَم

Muhammad the favourite and choices from all His creations.
One who is the last of the messengers of the Merciful God to the nations (the translation lacks starting each word with the m consonant).

9.2 *Distribution Characters/Alliteration 2* (توزيع 2)

This type distributes the same consonant or vowel in every word or most words of the entire poem. The programme checks, and we choose to record how many times the top three letters appear.

10. *Hemming* (تصريع)

This is when the end of the first hemistich matches the end of the verse in meter and final vowel. The following verse is by Abū Nuwās (d. 814 C.E.)

دَعْ عَنْكَ لَوْمِى فأنّ اللّوْمَ إغْراءُ ... وَداوِنى بالتى كانتْ هِىَ الدّاءُ

Censure me not, for censure but **tempts me** (*ighrā'u*); cure me rather with the cause of my **ill** (*al-dā'u*).

11. *Unraveling* (التوشيح)

This is use of a noun in dual form near the end of the second hemistich, followed by two single words specifying what constitutes this duality, the second being the rhyme word. An example from Mayyās al-Mawṣilī (no date of death is given):

أبيتُ فى لجَج التَذْكار مِنْكَ وبى ... حالان مخْتَلفان اليَأسُ والأمَل
لاَ يهَتدي لى طَيفٌ مُذْ هجرْتَ ولا ... يَزورُنى المُسلِيان الكَتْبُ والرُّسْلُ
أسائِلُ الدّارَ منْ وَجْدِى عليْكَ فلا ... بُجيبُى المُقفِران الرّبْعُ والطلَلُ

I spend the night importuned by memories of you, and in me
Are **two opposite conditions: despair and hope**.
No phantom finds its way to me since you deserted me,
Nor am I visited by the **two comforters: letters and messengers**.
My passion for you makes me question the site where I have halted, but
I get no answer from **either of the two desolate spots: the old encampment and its half-obliterated traces**.

12. *Counterchange* (عكس وتبديل)

This is repetition of words in a different order. It is of two kinds: (1) a hemistich is recast to complete the verse without any change in meaning, and (2) a hemistich or part thereof is re-arranged to produce a different meaning.

يا بَدَني يا بَدَني بالفِراق مُتْ كمدا مُتْ ... كَمَدًا بالفِراق

فارَقَني مَنْ أُحبُّ واحزَني ... واحزَني مَنْ أُحبُّ فارَقَني

Body of mine, now we are parted, die of grief!

Die of grief now we are parted, o body of mine!

He has left me, whom I love – O wretchedness!

Wretchedness, he whom I love has left me!

13. *Repartee* (مراجعة)

This is a reported conversation, combining pithy expression, subtle ideas, elegant composition, and flowing diction. This example is from the poetry of the Egyptian Ibn Maṭrūḥ (d. 1251 C.E.):

سَألتُ مَنْ أمْرَضَني ... في قُبْلةٍ تشْفي الأَلَمْ

فقال لا لا أبَدًا ... قُلتُ له نَعَمْ نَعَمْ

I asked of him who caused my languor for a kiss to ease the pain.

He answered, "No, no, never!" (*faqāla: lā lā abadan*)

I insisted, "Yes, oh yes!" (*qultu lahu: naʿam naʿam*)

14. *Rhyming – in general – homoeoteleuton* (سجع)

When the discourse is divided into sections with similar-sounding last syllables, it is said to be:

14.1. *Congruent* موازَن when the final words agree in measure as well as in rhyme, or

14.2. *Terminal* مطرف when they have rhyme but not measure in common. An example of congruent rhyming from the Qur`ān 88:13-14:

فيها سررٌ مرفوعة وأكوابٌ موضوعة

In it are thrones that **are raised** (*marfūʿa*), and goblets **placed ready** (*mawḍūʿa*).

An example of terminal rhyming from al-Waʾwāʾ al-Dimashqī (d. 995 C.E.):

قُمْ يا غُلامُ إلى المُدام ... قُمْ داوني منها يجام

Up, boy! To **the wine** (*al-mudām*), Medicate me with **a silver goblet** (*bi-jām*)

14.3. *Tucking* تشطير. The division of each hemistich into rhyming parts, but with different rhymes in each hemistich (producing the arrangement bbaa, where "a" is the rhyme common to all verses of the poem). An example from Abū Tammām in praise of the caliph al-Muʿtaşim bi-llāh (re. 833-842 C.E.), playing on the literal meaning of the honorific title:

تَدْبيرُ مُعْتَصِم باللہ مُنْتَقِم ... للہ مُرْتقَب في اللہ مُرْتغِب

He revenges [his enemies by the help of] God (*bi-l-lāhi muntaqimin*), and acts while finding refuge in God (*tadbīru muʿtaṣimin*)

He places his hope in God (*fī l-lāhi murtaghibin*), waiting to meeting Him (*li-lāhi murtaqibin*)

14.4. *Embroidery* ترصيع – Making every word in a hemistich or prose versicle agree in rhyme and possibly in measure, as well as in grammatical case with the correspondingly placed word in the next hemistich or versicle [Qur`ān 82: 13]:

إنّ الأبرارَ لَفي نَعيم, وإنّ الفُجَّارَ لَفي جَحيم

The (*inna*) righteous (*abrāra*) are in (*la-fī*) bliss (*naʿīmin*), and the (*wa-inna*) ungodly (*al-fujjāra*) are in (*la-fī*) hell (*jaḥīmin*)

15. *Verbal Congruence* (مناسبة لفظيّة)

Verbal congruence is the use of two words or sets of words cast in the same metrical pattern. It is of two varieties: perfect and imperfect. In the first, congruent words or versicles rhyme with one another. In the second, they do not. We focus on one of them: the verbal congruence perfect. An example of perfect verbal congruence, *nūru l-ghayāhibi* ("the light of darkness") and *jammu l-mawāhibi* ("abounding in gifts") from ʿAbd al-Ghanī al-Nābulsī:

نُورُ الغَياهِب في يَوْم الوَغى بَطَلٌ ... جمُّ المواهِب بَحْرُ الجُودِ والكَرَم

He is a light in darkness (*nūru l-ghayāhibī*), a hero on the day of strife,

Abounding in gifts (*jammu l-mawāhibī*), a sea of bounty and generosity

16. *Paronomasia/Pun* (جناس)

This is use or suggestion in one context of words that differ in meaning but are phonetically and/or graphically alike or nearly alike. They are of seven main types:

16.1. *Conjunct Paronomasia* (جناس مركّب - مَرْفُوّ)

A paronomasia in which a single word is matched by a combination of more than one. An example from Ibn Nubāta (d. 1366 C.E.), in which *amradā*, "beardless youth" , is matched by *am radā*, "or death:"

قمرًا نراهُ أمْ مليحا أمْرَدَا ... ولحاظُه بينَ الجوانح أمْ رَدَى

Is it a moon that we see, or a handsome **beardless youth** (*amradā*)?

Is it his glances [wreaking havoc] in my breast, **or is it death** (*am radā*)?

16.2. *Tipped Paronomasia* (جناس ناقص - مُطَرَّف)

A paronomasia in which one of the matching terms is longer than the other by at least one initial consonant. An example from Ibn al-Fāriḍ (d. 1235 C.E.):

إنْ كانَ فِراقُنا مع الصّبْح بَدا لا أسْفَرَ بَعْدَ ذاكَ صُبْحٌ ... أبَدا

If with the dawn your parting must **occur** (*badā*),

Then may no dawn **hereafter ever** (*abadā*) break

16.3. *Tailed Paronomasia* (جناس ناقص - مُذَيَّل)

A paronomasia in which one of the matching terms is longer than the other by one or more terminal consonants. An example by Ḥassān ibn Thābit (d. 674 C.E.):

وكّنا متى يَغْزُ النَبِيُّ قَبيلَةً ... نَصِلْ جانِبيها بالقنا والقنابل

Whenever the Prophet attacked a tribe,

We struck both flanks with **lances** (*qanā*) and with squadrons of **horse** (*qanābil*)

16.4. *Consonantal Paronomasia* (جناس مُحَرَّف; also *Distorted Paronomasia*)

A paronomasia in which the two matching terms consist of the same letters in the same order, differing only in vocalization. An example from Sharaf al-Dīn al-Anṣārī (d. 1264 C.E.):

لِعَيْنِي كُلَّ يَوْمٍ أَلْفُ عَبْرَه ... تُصَيِّرُنِي لأَهْلِ العِشْقِ عِبْرَه

My eyes shed every day a thousand **tears** (*'abra*)

So that to my lovers I have become **an object lesson** (*'ibra*)

16.5. Substitutive and Variant Paronomasia (جناس تصريف)

A paronomasia in which one of the matching terms has one letter – initial, medial or terminal – which differs from the corresponding letter in the other term, or is similar or close to their points of emission but does not differ in its dots (so that it belongs to a different or the same phonetic category). An example from Abū Firās al-Ḥamdānī (d. 968 C.E.):

تَعِسَ الحَريصُ وَقَلَّ ما يَأتي بِهِ ... عِوَضًا عن الإلْحاح وَالإلْحافِ

Wretched is the miser, and he seldom gets a return for his **insistence** (*ilḥāḥ*) and **importuning** (*ilḥāf*)

16.6. Metathetic Paronomasia (جناس مقلوب; also *Reverse Paronomasia*)

A paronomasia in which the two matching terms consist of the same letters, without addition or omission, but in different orders. An example from Shams al-Dīn Muḥammad al-Bakrī (d.1060 C.E.):

قلتُ مستعطفًا لساقٍ سقاني ... من طِلا نيل مِصْرَ أعذبَ كاسِ

أنتَ عندي أعزّ منه ولكنْ ... قَلْبُه لَيِّنٌ وقَلْبُكَ قاسِ

Pleadingly, I said to a **cupbearer** (*sāqī*) who poured out to me of the liquor of Egypt's **Nile** (*nīl*) a most limpid cup: You are dearer to me than it, and yet its heart is **soft** (*layyin*; the *y* in Arabic is written like the *ī*. Without the diacritics, the word *layyin* looks like *līn*), but yours is **cruel** (*qāsī*)

16.7. Graphic Paronomasia (جناس تصحيف)

A paronomasia such that, when written, the two matching terms are shaped alike, differing only in their dots (which, in Arabic script, differentiate many letters from one another). An example:

إنْ كان شرْعُ هواك أطْلَقَ أدمُعي ... فوَكيلُ شوْقي عاجزٌ عن حَبْسِه

أوْ كان منكَ الطَّرْفُ أسْهَرَ ناظِري ... فلِكُلّ شَيْءٍ آفَةٌ مِنْ جِنْسِه

Though the law of your love set free my tears,

Yet is the advocate of my yearning incapable of **restraining it** (*ḥabsihi* حبسه)

And, if your eyes deprive mine of sleep

[It is admitted that] all things are vulnerable to **kindred ill** (*jinsihi* جنسه)

17. Negative Antithesis (طباق سلب)

Combining the object (word) and its opposite in Arabic literature produces two main kinds of antithesis — positive and negative. Our focus was the negative, which is considered a type of Repetition, with the opposite word preceded by a negation article {لا, ما, لم, لن, ليس, لات, غير}. This example is from the poem by al-Buḥturī (d. 898 C.E.) between the phrases *knowing* and *not knowing*:

يقيّض لي من حيث لا أعلم النوى ... ويسري إليّ الشوق من حيث أعلم

Separation is predestined to me from a source that **I do not know** (*lā aʿlamu*);

But I feel the longing of a source that **I do know** (*aʿlamu*).

18. Positive Antithesis/Oxymoron (طباق إيجاب)

This is when words and their opposites appear in the same verse. Opposites formed by negation words are not positive antitheses. An example from Ibn Nubāta:

دعوتُ ألْفاظ المَليح وكأسَه ... فنعمْتُ بينَ حديثِه وعَتيقِه

I call forth the fair one's conversation and his cup, and so delight in his **present discourse** (*ḥadīthihi*) and **old** wine (*qadīmihi*).

19. The Satirist's Feint (تأكيد الذَّم بما يشبه المدح)

This either (a) denies that the subject has any praiseworthy quality, or (b) ascribes to the subject a pejorative attribute then, after exempting him/her from it or rectifying it, names another pejorative attribute. An example from al-Nābulsī:

فإن مَنْ لامَني لا خيرَ فيه سِوَى ... وصفي له بأخس الناس كُلِّهِم

My reprover **has no good qualities except** (*lā khayra fīhi siwā*)

being the vilest among all people (*akhassa l-nāsi kullihimi*).

20. The Encomiast's Feint (تأكيد المدح بما يشبه الذَّم)

This either (a) denies that the subject has any pejorative attributes, then exempts him/her from that denial, as if reversing it, a praiseworthy quality. Or (b) ascribes to the subject a praiseworthy quality then, after an exempting or rectifying word, names another praiseworthy quality. The following example is from the pre-Islamic poet an-Nābigha l-Dhubyānī (d. ca. 604 C.E.):

ولا عَيْبَ فيهم غيرَ أنّ سُيوفَهُمْ ... بِهِنّ فلولٌ مِنْ قِراع الكَتائبِ

There is no fault in them, excepting that their swords — have

Suffered dents in clashes with battalions.

Appendix 2: Poets and Number of Poems

Name of Poet	Number of poems in training set	Number of pomes in text set
Al-Buḥturī	3	
Ibn al-Rūmī	7	1
Abū Tammām	9	1
Al-Mutanabbī	2	
Dīk al-Jinn al-Ḥimṣī	10	
Al-Najāshī al-Ḥārithī	2	
Al-Kumayt ibn Zayd	2	
Ibn Qys al-Ruqayyāt	2	
Al-Ṭirimmāḥ b. Ḥakīm	2	
Jarīr	2	
Al-Farazdaq	3	
Abū Dhuʾayb al-Hudhalī	2	
Bashshār b. Burd	1	
Al-Aʿshā Maymūn	1	1
Imruʾ al-Qays	1	
Mulayḥ b. al-Ḥakam	1	
Abū Arāka	1	
Ayman b. Khuraym	1	
Taʾabbaṭa Sharran	1	
Dhū l-Rumma	1	
ʿUbaydallāh b. al-Ḥurr	1	
ʿImrān b. Ḥiṭṭān	1	
Qaṭarī b. al-Fujāʾa	1	
Yazīd b. Mufarrigh al-Ḥimyarī	1	
Al-Nābigha l-Dhubyānī		1
Rabīʿa b. Maqrūm		1
Al-Namir b. Tawlab		1
Al-Ḥakam b. ʿAbdal		1
Abū Dulāma		2
Anonymous (Abbasid poet)		1
Al-Ḥamdawī		3
Abū Nuwās		1
Muḥammad b. Yasīr al-Riyāshī		1
Abān al-Lāḥiqī		1
ʿAbdallāh b. ʿAbd al-Ḥamīd al-Lāḥiqī		1
Total count	**58**	**17**

Table 4.

Appendix 3: Number of Rhetorical Elements in the Training and Test Datasets

Elements	Appearances in the sets	Training set	Test set
Simile تشبيه	290	37	
Repetition1 (تكرار)	196	31	
Repetition2 (اشتقاق)	247	40	
Repetition3 echo (رد الاعجاز على الصدور)	131	17	
Repetition4 – Key Words كلمات مفتاح	66	45	
Flowing (اطراد)	5	3	
Afterthought (رجوع)	5	5	
Catchword Verbal (تشابه الاطراف اللفظي)	4	8	
Distribution Characters1/Alliteration 1 (توزيع 1)	5	5	
Distribution Characters2/ Alliteration2 (توزيع 2)	174	51	
Hemming (تصريع)	5	5	
Unraveling (توشيع)	16	14	
Counterchange (عكس وتبديل)	8	5	
Repartee (مراجعة)	6	5	
Rhyming – in general – homoeoteleuton (سجع)	5	24	
Verbal Congruence (مناسبة لفظيّة)	4	7	
Paronomasia (جناس)	32	21	
Negative Antithesis (طباق سلب)	66	7	
Positive Anthithesis / Oxymoron (طباق إيجاب)	245	35	
Satirist's, Encomiast's Feints تأكيد الذّم بما يشبه المدح) (تأكيد المدح بما يشبه الذّم	10	13	

Table 5.

Appendix 4: Appearance of Each Rhetorical Element in Each Poem (Rhetorical Elements are numbered 1,2,3,4,5,9.2,12,16,17,18

Poet	Poem number	Ele.1	Ele.2	Ele.3	Ele.4	Ele.5	Ele.9.2	Ele.12	Ele.16	Ele.17	Ele.18

Al-Buḥturī	51	8	5	4	7	0	3	0	0	3	5
	256	11	3	2	0	0	3	0	0	2	3
	915	10	4	15	3	0	3	1	1	1	11
Ibn al-Rūmī	9	8	2	9	3	6	3	1	0	1	6
	104	4	4	2	2	0	3	1	2	0	3
	343	4	2	2	1	0	3	0	0	1	3
	344	3	3	3	1	0	3	0	0	0	4
	539	0	2	2	4	0	3	0	0	2	8
	691	1	1	2	3	0	3	0	0	1	1
	701	3	5	2	7	0	3	2	0	0	2
Abū Tammām	2	10	6	6	3	0	3	0	2	1	6
	5	13	12	11	14	0	3	0	0	14	9
	21	2	0	2	4	0	3	0	0	0	3
	71	13	6	7	4	0	3	1	0	0	12
	179	9	5	1	4	0	3	0	0	2	5
	393	6	3	6	5	0	3	0	1	4	10
	464	1	2	2	0	0	3	0	0	4	8
	470	3	6	5	0	0	3	0	0	0	3
	484	4	2	6	1	2	3	0	0	3	10
al-Mutanabbī	59	21	10	7	1	0	3	0	1	1	9
	209	5	9	6	2	0	3	0	0	2	17
Dīk al-Jinn al-Ḥimṣī-	13	0	3	2	2	0	3	2	0	0	2
	14	0	1	0	0	1	3	0	0	2	1
	9	9	6	7	6	0	3	0	0	3	5
	62	4	5	3	3	0	3	1	0	0	0
	11	5	2	6	3	0	3	0	0	0	7
	15	0	3	1	1	0	3	0	0	0	1
	2	0	4	7	2	1	3	0	0	1	3
	35	0	0	0	0	1	3	0	0	0	0
	38	0	8	3	1	1	3	0	0	1	2
	33	1	1	1	0	0	3	0	0	0	0
Al-Najāshī al-Ḥārithī	41	4	1	1	1	0	3	0	0	0	1
	62	14	4	7	1	1	3	0	0	1	3
Al-Kumayt ibn Zayd	6	5	3	4	2	0	3	0	0	0	5
	8	1	2	3	0	0	3	0	0	0	0
Ibn Qys al-Ruqayyāt	40	2	0	2	0	7	3	0	0	0	2
	50	1	4	4	2	0	3	0	0	0	3
Al-Ṭirimmāḥ b. Ḥakīm	22	3	1	3	1	1	3	0	0	0	1
	290	1	1	3	1	3	3	0	1	1	3
Jarīr	3	4	9	3	5	2	3	0	1	4	7
	11	1	3	7	0	4	3	0	0	0	1
Al-Farazdaq	35	5	0	3	1	4	3	0	0	0	6
	50	0	2	3	0	2	3	0	0	1	5
	181	7	4	7	1	9	3	0	0	2	8
Abū Dhuʾayb al-Hudhalī	1	12	1	7	2	0	3	0	0	0	0
	2	4	2	5	1	7	3	0	0	1	6
Bashshār b. Burd	26	14	6	13	6	0	3	2	0	3	11
Al-Aʿshā Maymūn	79	13	1	2	1	0	3	0	0	1	3
Imruʾ al-Qays	8	8	1	4	2	0	3	0	0	0	0
Mulayḥ b. al-Ḥakam	2	12	3	2	1	7	3	0	0	0	2
Abū Arāka	1	0	0	0	0	0	3	0	0	0	0
Ayman b. Khuraym	17	1	2	3	2	1	3	0	0	0	1
Taʾabbaṭa Sharran	29	7	8	9	6	0	3	1	2	1	9
Dhū l-Rumma	30	10	0	5	0	0	3	0	0	0	1
ʿUbaydallāh b. al-Ḥurr	53	1	1	5	2	0	3	0	0	0	1
ʿImrān b. Ḥiṭṭān	199	1	1	4	0	6	3	0	0	1	6
Qaṭarī b. al-Fujāʾa	121	0	3	3	1	0	3	0	0	0	1
Yazīd b. Mufarrigh al-Ḥimyarī	44	1	8	3	5	0	3	0	0	1	0
Tally of poems in training set	**58**	**290**	**196**	**247**	**131**	**66**	**174**	**12**	**11**	**66**	**245**
Al-Nābigha l-Dhubyānī-	36	8	4	11	3	6	3	1	0	3	5
Rabīʿa b. Maqrūm	43	3	1	2	0	3	3	0	0	1	0
Al-Aʿshā Maymūn	78	6	4	5	2	1	3	0	0	0	5
Al-Namir b. Tawlab	34	2	3	3	0	2	3	0	0	1	1
Al-Ḥakam b. ʿAbdal	1	2	4	2	1	8	3	1	0	0	0
Abū Dulāma	1	4	0	1	1	0	3	0	0	0	0
Anonymous poet	9	0	0	0	0	0	3	0	0	0	1
Abū Tammām	338	0	1	3	0	4	3	0	0	1	3
Al-Ḥamdawī	12	0	1	0	0	4	3	0	0	0	0
Abū Nuwās	2	3	1	0	1	1	3	0	0	0	2

Muḥammad . Yasīr al-Riyāshī		2	2	2	0	1	3	0	0	0	1
Abū Dulāma	1	2	4	3	4	10	3	0	0	1	8
Ibn al-Rūmī	6	2	1	1	1	1	3	0	0	0	2
Al-Ḥamdawī	6	0	0	1	0	1	3	0	0	0	3
Al-Ḥamdawī	8	0	0	3	0	2	3	0	0	0	1
ʿAbdallāh b. ʿAbd al-Ḥamīd al-Lāḥiqī	4	1	3	1	1	0	3	0	0	0	2
Abān al-Lāḥiqī	1	2	2	2	3	1	3	0	0	0	1
Tally of poems in testing set	**10**	**37**	**31**	**40**	**17**	**29**	**51**	**2**	**0**	**7**	**35**

Table 6.

Acknowledgement

63

Notes

[1] https://arabic-rhetoric.haifa.ac.il/

[2] https://benjamins.com/online/met/

[3] https://arabic-rhetoric.haifa.ac.il/

[4] https://arabic-rhetoric.haifa.ac.il/ – link to the system

[5] https://ar.wikipedia.org/wiki/%D9%82%D8%A7%D8%A6%D9%85%D8%A9_%D8%A7%D9%84%D8%A7%D9%86%D9%81%D8%B9%D8%A7%D9%84%D8%A7%D8%AA_%D8%A7%D9%84%D8%B9%D8%A7%D8%B7%D9%81%D9%8A%D8%A9

Works Cited

Abdul-Mageed, Diab, and Kübler 2013 Abdul-Mageed, M., Diab, M.T., and Kübler, S. (2013) "ASMA: A System for Automatic Segmentation and Morpho-Syntactic Disambiguation of Modern Standard Arabic", *RANLP*. https://aclanthology.org/R13-1001/.

Abuata and Al-Omari 2018 Abuata, B., and Al-Omari, A. (2018) "A rule-based algorithm for the detection of Arud meter in classical Arabic poetry", *International Arab Journal of Information Technology*. 15(4):661-667.

Abushihab 2020 Abushihab, Ibrahim. (2020) "A stylistic analysis of Arab-American poetry: Mahjar (place of emigration) poetry", *Journal of Language Teaching and Research* 11(4):652-661.

Ahmed and Trausan-Matu 2017 Ahmed, M. A., and Trausan-Matu, S. (2017) "Using natural language processing for analyzing Arabic poetry rhythm", *16th RoEduNet Conference: Networking in Education and Research (RoEduNet)*. IEEE.

Al-Jāḥiẓ 2019 Al-Jāḥiẓ, ʿAmr b. Baḥr. (2019) *Al-Maḥāsin wa-l-aḍdaād*. Cairo: Muʾassasat Hindāwī.

Al-Jārim and Amīn 1999 Al-Jārim, ʿAlī, and Amīn, Muṣṭafā. (1999) *Al-Balāgha al-wāḍiḥa*. Cairo: Dār al-Maʿārif.

Al-Khaṭīb 2003 Al-Khaṭīb, al-Qazwīnī. (2003) *Al-Īḍāḥ fī ʿulūm al-balāgha*. Ed. Ibrāhīm Shams al-Dīn. Beirut, Dār al-Kutub al-ʿIlmiyya.

Alkhatib and Shaalan 2017 Alkhatib, Manar, and Khaled Shaalan. (2017) "Natural language processing for Arabic metaphors: a conceptual approach", *Advances in Intelligent Systems and Computing* 533: 170-181.

Badaro et al. 2014 [8] Badaro, Gilbert, et al. (2014) "A large-scale Arabic sentiment lexicon for Arabic opinion mining", *ANLP* 2014 165.

Boudchiche et al. 2017 Boudchiche, Mohamed, et al. (2017) "AlKhalil morpho sys 2: A robust Arabic morpho-syntactic analyzer", *Journal of King Saud University-Computer and Information Sciences*. 29(2):141-146.

Buckwalter 2002a Buckwalter, Tim. (2002) *QAMUS: Arabic Lexicography*. http://www.qamus.org/morphology.htm

Buckwalter 2002b Buckwalter, Tim. (2002) *Buckwalter Arabic Morphological Analyzer Version 1.0*. "LDC2002L49". Web Download. Philadelphia: Linguistic Data Consortium.

Cachia 1998 Cachia, Pierre. (1998) "The Arch Rhetorician or the Schemer's Skimmer", *A Handbook of Late Arabic Badīʿ Drawn from ʿAbd al-Ghanī al-Nābulsī's Nafaḥāt al-azhār ʿalā nasamāt al-asḥār*. Wiesbaden: Otto Harrassowitz Verlag.

Diehl et al. 2009 Diehl, Frank, et al. (2009) "Morphological analysis and decomposition for Arabic speech-to-text systems", *Tenth Annual Conference of the International Speech Communication Association*. Brighton, United Kingdom.

ElSahar and El-Beltagy 2015 ElSahar, Hady, and Samhaa R. El-Beltagy. (2015) "Building large Arabic multi-domain resources for sentiment analysis", *CICLing* (2).

Elkateb et al. 2006 Elkateb, Sabri, et al. (2006) "Building a WordNet for Arabic", *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, Genoa, Italy.

Englisch-hilfen.de n.d. Englisch-hilfen.de. n.d. *Opposites in English — Word list — A-F*. https://www.englisch-hilfen.de/en/words/opposites1.htm

Fayyūd 1998 Fayyūd, ʿAbd al-Fattāḥ Basyūnī. (1998) *ʿIlm al-bayān*. Cairo: Muʾassasat al-Mukhtār.

Fellbaum and Alkhalifa 2006 Alkhalifa, C., M., Black, W., Elkateb, S., Pease, A., Rodriguez, H., and Vossen, P. T. J. M. (2006). "Introducing the Arabic wordnet project." In *Third Global Wordnet Conference.*

Habash and Rambow 2006 Habash, Nizar and Owen Rambow. (2006) MAGEAD: "A morphological analyzer and generator for the Arabic dialects", *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics. Sydney, Australia.

Habash, Rambow, and Roth 2009 Habash, N., Rambow, O., and Roth, R. (2009) "MADA+ TOKAN: A toolkit for Arabic tokenization, diacritization, morphological disambiguation, POS tagging, stemming, and lemmatization", *Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR)*, Cairo, Egypt. Vol. 41.

Hayes-Roth 1985 Hayes-Roth, F. (1985) "Rule-based systems" *Communications of the ACM* 28(9): 921-932.

Heinrichs 1994 Heinrichs, Wolfhart. (1994) "Muslim b. al-Walīd und badīʿ" in Wolfhart Heinrichs and Gregor Schoeler (eds.) *Festschrift Ewald Wagner zum 65. Geburtstag*. Beirut: In Kommission bei Franz Steiner Verlag Stuttgart: 211-245.

Hussein 2015a Hussein, Ali Ahmad. (2015) "The rhetoric of Hudhalī wine poetry", *Oriens* 43(1-2):1-53.

Hussein 2015b Hussein, Ali Ahmad. (2015) *The Rhetorical Fabric of the Traditional Arabic Qaṣīda in its Formative Stages: A Comparative Study of the Rhetoric in Two Traditional Poems by ʿAlqama l-Faḥl and Bashshār b. Burd*. Wiesbaden: Harrassowitz Verlag.

Hussein 2018 Hussein, Ali Ahmad. (2018) "Majāzʿ aqlī 'intellectual trope' and the description of wine in a poem by Abū Dhuʾayb al-Hudhalī", *Acta Orientalia Academiae Scientiarum Hungaricae*. 71(4):429-442.

Hussein n.d. Hussein, Ali Ahmad. No date. "Statistical development of the use of rhetorical elements in classical Arabic poetry", unpublished.

Maamouri et al. 2010 Maamouri, Mohamed, et al. (2010) "Standard Arabic morphological analyzer (SAMA) version 3.1", *Linguistic Data Consortium*. Catalog No. LDC2010L01.

Miller 1995 Miller, George A. 1995. "WordNet: a lexical database for English", *Communications of the ACM*. 38(11):39-41.

Moreau 2003 Moreau, François. (2003) *Al-Balāgha. Al-Madkhal li-dirāsat al-ṣuwar al-bayāniyya*. Trans. Muḥammad al-Walī and ʾĀ ʾisha Jarīr. Morocco and Beirut: Afrīqyā l-Sharq,

Pasha et al. 2014 Pasha, Arfath, et al. (2014) "MADAMIRA: A fast, comprehensive tool for morphological analysis and disambiguation of Arabic" *LREC*. Vol. 14.

Perfectyourenglish.com 2019 Perfectyourenglish.com. (2019) *100 Opposite Words*. http://www.perfectyourenglish.com/vocabulary/opposite-words.htm

Sasaki 2007 Sasaki, Yutaka. 2007. "The truth of the F-measure", *Teach Tutor Mater*. 1(5):1-5.

Sawalha, Atwell, and Abushariah 2013 Sawalha, Majdi, Eric Atwell, and Mohammad AM Abushariah. (2013) "SALMA: standard Arabic language morphological analysis. Communications, Signal Processing, and their Applications (ICCSPA)", *1st International Conference on Communications, Signal Processing, and their Applications IEEE*. Sharjah, United Arab Emirates.

Schaade and von Grunebaum 2012 Schaade, A., and G.E. von Grunebaum. (2012) "Balāgha", *The Encyclopaedia of Islam2*. Online version. 1: 981-982.

Shoukry 2013 Shoukry, Amira Magdy. 2013. *Arabic sentence-level sentiment analysis*. PhD thesis, the American University in Cairo, Cairo.