

ARTICLE

The Word Formation Process between the Arabic and Indonesian Language: Corpus Linguistics in Morphological Approach

Muhammad Yunus Anis^{ID}

Arabic Literature Department, Universitas Sebelas Maret, Surakarta 57126, Indonesia

ABSTRACT

This article examines the word formation process of Arabic and *Bahasa Indonesia* comparatively from the point of view of corpus linguistics in morphological approach. Arabic as one of the Semitic language families has a distinctive character in word formation. This can be seen in the process of word formation through five important stages, namely: (1) inflection, (2) derivation, (3) allomorphy, (4) abbreviations - acronyms, and (5) compounding. These five things play an important role in the process of word formation in Arabic. In general, this research design is qualitative research. The data in this study are representative language units to examine word formation in Arabic and Indonesian. Then corpus linguistics becomes one of the media in validating the data. The results of the discussion of the article are expected to provide new insights into Arabic, especially related to the study of Arabic Morphology (*Sharaf Science*) based on corpus linguistics approach. After a comparative analysis based on corpus linguistics and morphology, it can be concluded that, some words in the form of stems (*shighatul asās*), after experiencing the inflection process, the stem-shaped words have decreased the number of frequencies in the corpora. The study of the morphological word formation process based on corpus linguistics has a contribution to the preparation of lexical dictionaries and language planning, both Arabic and Indonesian. This research is only limited to the study of morphology, word formation, in the future, the potential for corpus studies can be expanded from the side of syntax and semantics.

Keywords: Inflection; Derivation; Allomorphy; Abbreviations; Acronyms; Compounding; Morphology; Corpus Linguistics

*CORRESPONDING AUTHOR:

Muhammad Yunus Anis, Arabic Literature Department, Universitas Sebelas Maret, Surakarta 57126, Indonesia; Email: yunus_678@staff.uns.ac.id

ARTICLE INFO

Received: 4 April 2025 | Revised: 30 April 2025 | Accepted: 9 May 2025 | Published Online: 19 May 2025

DOI: <https://doi.org/10.30564/fls.v7i5.9360>

CITATION

Anis, M.Y., 2025. The Word Formation Process Between the Arabic and Indonesian Language: Corpus Linguistics in Morphological Approach. *Forum for Linguistic Studies*. 7(5): 1112–1127. DOI: <https://doi.org/10.30564/fls.v7i5.9360>

COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

1. Introduction

The morphological analysis of Arabic is very dominant with the analysis of form and meaning, there are still very few studies that specifically discuss the relationship between Arabic morphology and corpus linguistic approaches. Several experts have defined morphology as the process of word formation. Morphology is: (1) the field of linguistics that studies morphemes and their combinations, in this first definition, the study of morphology is specifically focused on the study of “morphemes” of a language; (2) part of the language structure that includes words and parts of words, namely morphemes, meanwhile, the second definition specializes more in the study of morphology on “words” and their parts, one of which is “morphemes”^[1].

From these two definitions, it can be concluded that in essence the study of morphology is centered on words and morphemes. In some references, the term morphology in Arabic is often equated with the term */ilmu sharfi/* or */sharfiyyun/* (علم الصرف - صرفي), namely: the study of morpheme and their different forms (allomorphs) and the way they combine in word formation^[2]. From this definition, it can be concluded once again that the study of morphology is centered on morphemes and several different morpheme forms, namely allomorphs. Then morphology is also in charge of composing word formation. Therefore, this article will discuss the two basic principles of word formation (borrowing Katamba’s term: two principal word-building processes)^[3]. In this case, the study will focus on the discussion of morphology and its interaction with syntax. Broadly speaking, the discussion will focus on five main sub-chapters, namely: (1) inflection, (2) derivation, (3) allomorphy, (4) abbreviations – acronyms, and (5) compounding. These five things play an important role in the process of word formation in Arabic and Bahasa Indonesia based on corpus linguistics in morphological approach. The selection of Arabic and Indonesian as the main object of study in this comparative research is due to several reasons, such as: (1) the existing word formation studies have been carried out partially^[4,5], studies that specifically examine the comparative analysis of corpus-based word formation are still quite few^[6], (2) Arabic has a very different formation process from Indonesian, in this case, Arabic uses a flexion pattern where the process or result of adding

affixes to the base or root to limit grammatical meaning, on the other hand, Indonesian uses the agglutinative type, Indonesian has a word structure and grammatical relations characterized by the free combination of elements^[1]. In addition to Arabic and Indonesian having differences, they also have similarities, such as: a lot of Arabic words used in Indonesian^[7]. Therefore, a comparative study between Arabic and Indonesian, especially in terms of corpus-based word formation, is important.

Studies related to morphology have been carried out previously by Erwin and Loe about the reduplication process and its rules in the Rote language^[8]. This study concludes that there are three kinds of reduplication, namely: (1) full reduplication, (2) partial reduplication, and (3) verb-specific full reduplication. The study of reduplication is essentially important to be discussed comparatively, especially between Arabic and Indonesian. This is because word formation in Arabic does not entirely arise from reduplication. The study of morphology in the Qur’an has been reviewed in the *Lingua* journal by Aunul Hakim^[9]. By paying attention to previous studies, this research seeks to complement existing studies regarding the comparison of Arabic and Bahasa Indonesia based on morphological studies and corpus linguistics. The study of Arabic morphology began to be developed from the side of corpus linguistics^[10]. A corpus is often defined as a collection of written or spoken utterances used to support or test hypotheses about language structure^[1]. Corpus linguistics is a linguistic method that uses data from language materials collected in a source (a kind of language “bank”) derived from language use in various genres, variations, and spoken and written materials that ensure the widest possible variety and that avoid very narrow language use such as idiolects; the data is systematically organized and usually easily accessible electronically (by computer). In Arabic, the term corpus is equivalent to many terms, such as *matan* (متن) and *mudawwanah* (مدونة)^[11].

This study focuses more on the morphemic addition process based on corpus linguistics. Thus, the study of inflection and derivation has not been studied in detail, as well as allomorphy, abbreviations, and compounding have also not been studied in detail by previous researchers. This can be one part of the research gaps to be developed into a novelty of study and innovation. To see the com-

parative study model of Arabic and Urdu reduplication, it can be read from a study that was previously conducted^[12]. From the study, it can be concluded that reduplication, especially the “reduplicative compound” model in Arabic can be found in verbal speech. In addition, the reduplicative compound can also be found in the phenomenon of reduplication itself and repetition. In particular, the study of word formation processes in Arabic and English has been done before^[13]. In this study, Jarf focuses on the relationship between the Arabic word formation process and translation studies^[13]. Some of the Arabic word formation processes elaborated by Jarf are (1) acronyms, (2) abbreviations, (3) word coinage, (4) blends, (5) back-formations, (6) shortening (clipping), (7) extension, (8) conversion, (9) onomatopoeia, (10) borrowings, and (11) periphrasis. Some things that have not been studied in this article are that the study conducted by Jarf does not use a comparative method, and the flow of discussion starts from English to Arabic. As for this study, what distinguishes it from the research done by Jarf is the flow of analysis that starts from Arabic to Indonesian by using a comparative study. Jarf defines an acronym as a word formed from the first letter of a word in a phrase, as an example of an acronym in English according to Jarf is USA (United States of America), then Jarf gives examples of acronyms in Arabic found in the Qur'an such as yasin (يس), thaha (طه), and ha'-mim (حم). What missing literature is the utilization of corpus engines to validate word formation process data in Arabic and Bahasa Indonesia. Meanwhile, this article will only discuss in more detail and comparatively related to five aspects of the word formation process in Arabic, namely: (1) inflection, (2) derivation, (3) allomorphy, (4) abbreviations - acronyms and (5) compounding. For other word formation processes, it is suggested to be investigated in future research in more detail, such as: four phrases, namely: *murakkab ismiy* (nominal phrase), *murakkab fi'liy* (verbal phrase), *tarkib idāfiy* (Combination of Nouns), *tarkib wasfiy* (adjective phrase)^[14].

Comparative studies between Palestinian dialects and Modern Standard Arabic have been conducted by previous researchers^[15]. The study discusses the morphological process in morphemes which in turn can see the morphological distance between written language and spoken language as the mother tongue of children. The distance

between the two language varieties morphologically affects the development of children's literacy acquisition. The morphological distance between Palestinian dialects and Modern Standard Arabic (MSA) is important to study. At least from this article, we are inspired to conduct a comparative morphological analysis. This study attempts to map this distance based on form and function in three dominant systems: derivation, inflection, and clitics on nouns and verbs. All morphemes are arranged on an axis targeting four main levels that vary in terms of morphological distance from (spoken Palestinian Dialect) SPD to MSA, starting with 'identical' morphemes (same form, same function), followed by 'highly overlapping' morphemes (morpho-phonological differences in form, same function), then 'partially overlapping' morphemes (different form, same function), and ending with 'unique' morphemes (unique form, same/different function). Implications for the possible impact of morphological distance on reading acquisition are discussed. However, this research on Arabic morphology has not specifically examined the use of corpus data engines. The corpus allows Arabic and Indonesian students studying a foreign language to be able to identify more possible contexts in which to respond to the word. Previous studies by other researchers show that corpus studies have not been specifically applied in the process of morphological word formation, whether in Arabic, Indonesian or other languages^[4,16-21]. The study of the corpus can also be linked beyond morphological aspects, such as: discourse aspects^[22], stylistic analysis^[23], linguistics for vocabulary^[24], teaching Verb+Noun collocations^[25], semantic prosody and collocation^[26], however, in this study, the corpus will only be applied specifically within the morphological paradigm. This is one of the main research gaps offered in this study.

Comparative studies between Arabic and English have also been carried out previously^[27], the research focuses on transformative grammar and the workflow is a comparison of Arabic and English. Therefore, the author tries to compile a comparative analysis study of Arabic and bahasa Indonesia by looking at some studies that have been done before in other languages. This research is expected to contribute to the “policy maker” process carried out by the Arabic language agency and the Indonesian language agency, especially in the study of word formation process-

es based on morphological analysis and corpus linguistics, especially in the process of compiling the dictionary.

2. Materials and Methods

The research method in this study in general is to use a qualitative descriptive model. The data is taken from the units of the language of standard Arabic. The data are words and morphemes derived from selected data sources in standard Arabic. Data selection in the study was carried out by taking one case example of the word formation process in Arabic and Indonesian. The data is generally taken from dictionaries and some corpus engines, such as: monolingual corpus and parallel corpora. This research is a synchronic or descriptive research, which is concerned with events that occur within a limited period and does not involve historical development^[1]. In this case, the research method is divided into three main steps, namely: (1) data collection, (2) data analysis, and (3) result.

Data collection is generally carried out using observation methods based on the process of word formation in Arabic and Indonesian. The analysis is carried out using the distributional method along with its elaboration techniques, namely the direct element division technique to determine the types and processes of word formation in Arabic and Bahasa Indonesia. The last stage to be done is the presentation of the results of data analysis. The results of the data analysis are presented in a report. The presentation of the report is presented descriptively, namely the presentation of the results of data analysis using ordinary words. In this presentation, formula(s) or rule(s) are presented using ordinary words^[28].

Specific steps to be taken in corpus analysis such as: (1) determination of the object of study, such as morphology, syntax, semantics, pragmatics, or discourse, (2) selection of language units to be analyzed and elaborated using corpora, such as: morphemes, words, and phrases, (3) determination of context, corpora machines are sufficient in mapping the language context. To strengthen data analysis in the process of Arabic and Bahasa Indonesia word formation, this study uses several corpus data sources, such as: monolingual corpus (korpusindonesia.kemdikbud.go.id), the Quranic Arabic Corpus, Corpora Collection Leipzig University, Tunisian Arabic Corpus), and parallel corpora (Leeds Corpus of Contemporary Arabic: powered by CQP-

web, The Qur'an (Arabic): powered by CQPweb, Arabic – English Parallel Corpus). Several corpus engines were used in the study to validate the data on which Arabic, English and Indonesian word formation is based. In this study, some words (language units) were selected based on the word formation process, such as: *Inflection*; *Derivation*; *Allomorphy*; *Abbreviations*; *Acronyms*; *Compounding*.

Linguistic patterns in this study are identified from the morphological character of words, such as patterns of inflection and derivation patterns in Indonesian and Arabic, which in both languages have different formation patterns, but the similarity is that we can identify the stem of both languages, such as: roots. The stem form and the result of stem formation into a derivative form can be compared by using a corpus machine. Meanwhile, the linguistic patterns related to *allomorphy* in Arabic and Indonesian are related to phonological cases. Different realizations of *allomorphy* cases can be traced through the corpus machine. In the end, the linguistic patterns found in *abbreviations*; *acronyms*, and *compounding* in Arabic and Indonesian are formed by compacting and combining word elements, in this case the corpus engine validates how the original word conditions before and after experiencing the abbreviations; acronyms and compounding process.

3. Results

In general, the results of this study show that Arabic and Indonesian have distinctive characters in the process of word formation, especially in morphological terms. Meanwhile, the analysis in corpus linguistics shows the vitality of a word, the root that is still pure, (has not received additions or affixation, or the process of combining words), has a high number, when the root undergoes the process of inflection, derivation, the root will experience a decrease in number. Likewise, in the case of allomorphy, compounding, abbreviations, and acronyms, in this case the corpus engine can review the decrease in the number of roots, base, and stem usage.

3.1. Inflection and Derivation

Inflection is (1) a change in word form that indicates various grammatical relations; including declination of

nouns, pronominal, and adjectives, and conjugation of verbs; (2) an element added to a word to indicate a grammatical relation; for example, *s* in *boys* indicates plural inflection, *s* in *reads* indicates third person verb inflection^[1]. Some linguists conclude that inflectional morphology has a study of word formation related to syntax. On the other hand, if we borrow the term from^[29], morphology is a study that discusses (word structure), both in terms of word structure organization, rules, and processes that focus on discussing meaningful units of language, which can be the “word” itself or part of the word, namely affixes. In this case, Katamba also distinguishes between inflectional morphology and derivational morphology. The first study, inflectional morphology, oversees determining the process of affixation related to syntax. Meanwhile, derivation morphology oversees forming of new lexical elements. The derivation is the process of affixing non-inflective affixes to a base to form a word^[1].

This article will specifically elaborate on the study of inflection and derivation using Arabic data. The study of Arabic morphology is certainly different from the study of English morphology, but it cannot be denied that theoretical studies (key concepts) in the development of theoretical studies of Arabic morphology have been influenced by the study of Semitic languages^[29]. The study of morphology which focuses on discussing word structure has two important issues, namely: (1) how words are formed (this is closely related to derivational studies or often referred to as lexical morphology) and (2) how they interact with syntax (this is related to inflectional morphology, such as the study of gender marker categories, number, case, and tenses). Arabic grammarians began to study Arabic morphology around the end of the eighth and ninth centuries after Christ. Ryding emphasized that the study of derivational or lexical morphology examines the basics of the formation of word formations (such as in English, the word *truthful* or *untruthfulness* comes from the derivation process of the word *true*).

Meanwhile, inflectional morphology explains how words have various forms, and how words undergo inflection to express grammatical categories such as singular/plural, or past/present tense. On the other hand, when “derivation” is positioned as a process of forming words or lexical units, then procedurally, derivation acts to precede

the process of inflection.

In this case, the derivation process occurs at the word root level and modifies it. Stems are morphemes, words or phrases that are joined by affixes; for example, in “*men-golah*” cultivate, “*bertani*” farm, “*pertanggungjawaban*” responsibility, the forms “*olah*” cultivate, “*tani*” farm and “*tanggung jawab*” responsibility are stems. This concept is distinguished from “base”; the form “*tanggung jawab*” responsibility is a single stem, but consists of two bases. Meanwhile, a “simple stem” is a stem that occurs from one basic morpheme, for example, “*tani*” in “*bertani*” farming^[1].

In the case of inflection, the word “*tani*” in Indonesian has a high frequency of 134 words in the category of regulations and laws. Then the word *tani* is found in 129 words in the biological category. After the word “*tani*” as the *stem* undergoes the inflection process to become “*bertani*”, the word has decreased in frequency in the corpus. The number of “*bertani*” words was found to be 25 in the social category and 22 in the cultural category.

The word (*tanggung jawab*) “responsibility” as a stem has a high number in the social category, around 963 data. After the word undergoes the inflection process (*tanggung jawab* –becomes *pertanggungjawaban*) or (responsibility - accountability), the word has decreased in number to 374 in the social category. Thus, it can be concluded that the stem has a high frequency compared to the word that has undergone the inflection process. The inflection process is proven to decrease the number of frequencies of word usage (source: korpusindonesia.kemdikbud.go.id).

In the Leipzig corpus, the word “*tani*” (number of occurrences: 26,026) and “*farming*” (number of occurrences: 4,995) were found in a total of 74,329,815 sentences containing the word “*tani*”. For example, in the following sentence:

Bagian yang amat besar dari kaum buruh dan tani sama-sekali tidak mempunyai hak pilih. (www.marxists.org, collected on 08/05/2012).

A very large section of the workers and peasants are completely disenfranchised.

Mata pencaharian warga dukuh ini selain bertani juga ada yang berdagang dan wiraswasta. (pnpmmpsukoharjo.blogspot.com, collected on 08/05/2012)

The livelihoods of the residents of this hamlet are not only farming, but also trading and self-employment.

The word graph of the two words can be seen in **Figure 1** and **Figure 2** below:

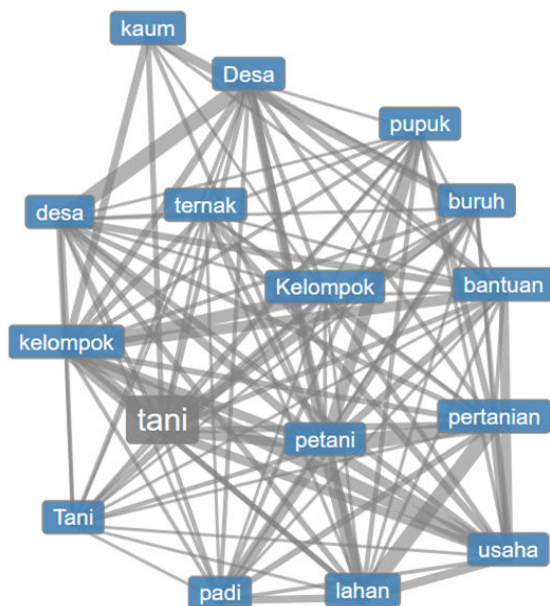


Figure 1. The Word Graph of “tani” (farm), kaum (the people) (peasantry), desa (village), ternak (cattle), pupuk (fertiliser), buruh (labour), kelompok (group), bantuan tani (farm aid), petani (farmer), pertanian (agriculture), padi (rice), lahan tani (farmland), usaha tani (farming business).

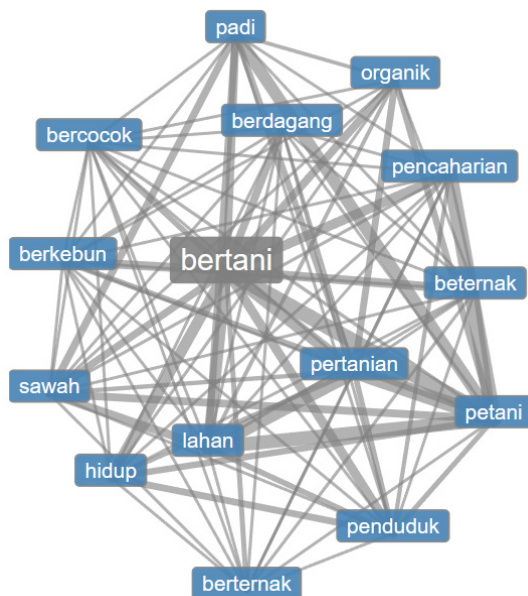


Figure 2. The Word Graph of “bertani” (farming), bertani padi (rice farming), bertani organik (organic farming), bercocok tani (farming), berdagang (agricultural trade), pencaharian bertani (farming livelihood), beternak (animal husbandry), sawah pertanian (rice field farming), lahan pertanian (agricultural land), hidup bertani (farm life), penduduk bertani (farming population), berternak (cattle), pertanian (agriculture), petani (farmer).

3.2. Allomorphy

An allomorph (allomorph, morpheme alternant) is a morpheme member whose position has been determined, e.g. [bær], [bə], [bəl] are allomorphs of the morpheme {ber-}^[1].

In the case of bahasa Indonesia, (me-) is a morpheme that has six allomorphs, namely (me-) itself, (men-), (meny-), (mem-), (meng-), and (menge-). The use of the first five allomorphs must be adjusted to the phoneme that follows it. The use of allomorphy cases in the corpus can be seen from the original word before getting additional allomorphs, such as the word “rawat” or “care”, in the Indonesian corpus appears around 1,197 in the health category. But after getting an additional morpheme (me), the word (rawat) becomes (merawat), its appearance in the corpus decreases to 481 in the same category, namely: health.

In Arabic, it is often equivalent to the term /shūrah sharfiyyah/ (صورة صرفية), ie: any of the different forms of a morpheme, for example, in English the plural morpheme is often shown in writing by adding –s to the end of a word, example: cat – cats /s/. Sometimes this plural morpheme is pronounced /z/, for example, dog – dogs /z/, and sometimes it is pronounced /iz/, for example, class – classes /iz/, all have the same grammatical function in these examples^[2].

In this study, allomorphs will intersect with the study of “assimilation” in Arabic. Sound assimilation occurs when several sounds become one and are pronounced in a double sound. The study of sound assimilation as an important part of Arabic allomorphs has been studied before (Al-hamadi & Salman, 2011)^[30]. There are three kinds of sound assimilation in Arabic, namely: (1) identical assimilation /idghām al-mutamāsilain/ (إدغام المتماثلين), (2) approximate assimilation /idghām al-mutaqāribain/ (إدغام المتقاربين), dan (3) /homogenous assimilation /idghām al-mutajānisain/ (إدغام المتجانسين).

In the case of allomorphs in Arabic, the *idgham* phenomenon in this study will be studied in the CQPWEB corpus, the 1,001 Nights (Arabic raw text) (<https://cqpweb.lancs.ac.uk/arabic1001>). One example of *idgham* phenomenon is the word (min) (من) and (māl) (مال). When we check in the corpus machine the word (māl) before getting the *idgham* phenomenon has a fairly high number position, but after we notice the *idgham* phenomenon in the word (min + māl) (من مال), the number of word frequencies

decreases. This can be seen in **Figure 3** (the word *māl* in CQPWEB) and **Figure 4** (*idgham* in CQPWEB).

Some explanations of the language units that appear in the **Figure 3** above. The word (مال) “fortune, riches, and welath” can be paired with the word (كثير) “a great deal, abundant, many, much”.

Based on the examples in the corpus listed in **Figure 4** above, it can be concluded that the word *min māl* (من مال) has quite a few uses, the word (يُخَرِّجُ) in the example above being the main basis in the formation of the phrase (التاجر) or “*min mālīt-tājir*” or “from the treasure of a merchant”.

Your query “مال” returned 101 matches in 1 text (in 597,628 words [1 text]; frequency: 169.001 instances per million words) [0.321 seconds]			
<	<<	>>	>
Show Page:	1	Line view	Show in random order
Choose action...			
No	Text	Solution 1 to 50	Page 1 / 3
1	1001	كثير وأولاد وزوجة وعندي رهون فدعني أذهب إلى بيتي وأعطي كل ذي	ساعة فقال التاجر للعفريت أني على دين ولي
2	1001	وجنود وبأس وأعوان من سائر الأجناس ، وكان في جسده برص قد	في مدينة الفرس وأرض رومان ملك يقال له الملك يوثان وكان ذا
3	1001	من مالي فقالنا لي أن الزواج خير لنا وليس لنا صبر عنه	سواء وأحسننا إليهما غاية الإحسان فبكنا عندي مدة سنة كاملة وصار لهما
4	1001	قلبي له ثم جاء وجلس وإذا بالقاضي قد دخل ومعه أربعة شهود	الحسن فوق وجنتيه أشهد أن لا ملج إلا هو فلما نظرت إليه
5	1001	كثير ، فلما كان آخر السنة جاء وعليه بدلة أفخر من الأولى	نفسى : والله إذا جاء لأضيفته لكوني انتفعت بدارهه وحصل لي منها
6	1001	وهي بنت أمير مات والدها وخلف لها مالا كثيرا . فودعته وانصرفت	الحب سألت التاجر عنها حين أردت القيام فقال : إن هذه صاحبة
7	1001	كثير فاشترى دارا حسنة واشترى المساليك والخيل والسروج المذهبة وأكل وأشرب ولا	زجاجا أبغعه بأربع مائة درهم ولا أزال أبغع وأشتري إلى أن يبقى عسى
8	1001	واتجر فيه وأترك اللهب واللعب . ثم إنه نهض من وقته وساعته	واروح إليهم وأطرق أبوابهم لعل أنال منهم شيئا فأجعله في يدي رأس
9	1001	وله ولد كاته البدر ليلة تسامه فصيح اللسان اسمه غاثم بن أيوب	أنه كان في قديم الزمان وسالف العصر والأوان تاجر من التجار له
10	1001	فلما راها غاثم بن أيوب عرف أنهم تغامزوا عليها ، فلما تحقق	وسماخ من الذهب وقلاند من الجوهر تساوي ملك السلطان ما بقي بثمنها
11	1001	إليهم وأقبل عليهم وسألهم عن حالهم وما سبب إقبالهم فقبلوا الأرض بين	والأ فلا مرد لأمره فعدت ذلك أمر لهم بالدخول فلما دخلوا عليه
12	1001	، والأخر فيه رجال يحفظون تلك الهدايا ممن يتعرض لها في البحر	هدية من التحف والمال ومن جعلتها ثلاث خرزات وجهاز مركبين واحد فيه
13	1001	أجود من العفل ولا عقل كالتدبير والحزم ولا حزم كالتقوى ولا قرية	يأكل من الفقه وأن لا يؤخر الصلاة عن وقتها وقيل : لا
14	1001	المسلمين فلا يا حفصة قد أرضيت قومك وأغضبت أبك فقامت تجر ذيلها	فقال : يا حفصة إنما أوصى الله بحق قرابتي من مالي وأما
15	1001	، قال : نعم إذا كان متى صبر ومتى أعطى شكر ؟	هو قادم عليه . وقيل لسفيان : هل يكون الرجل زاهد وله
16	1001	ولكن التقى هو السعيد وتقوى الله خير زاد حقا وعند الله تلقى	خير زاد في الميعاد كما قال بعضهم : ولست أرى السعادة جمع
17	1001	عن جواده فأسندوه وجعلوا به إلى أخيه ضوه المكان . ثم أوصوا	أسبق الناس إليه الوزير دندان وأمير الترك بهرام وأمير الديلم فلحقوه وقد
18	1001	وقماش يحضر لك سرعا ولا أكلفك بشيء أبدا وأيضاً عندي دائماً الخبز	أرديك إلا بسنة الله ورسوله صلى الله عليه وسلم وهما أردت من

Figure 3. The Word *māl* in CQPWEB.

Your query “من مال” returned 8 matches in 1 text (in 597,628 words [1 text]; frequency: 13.386 instances per million words) [0.437 seconds]			
<	<<	>>	>
Show Page:	1	Line view	Show in random order
Choose action...			
No	Text	Solution 1 to 8	Page 1 / 1
1	1001	وقماش يحضر لك سرعاً ولا أكلفك بشيء أبداً وأيضاً عندي دائماً الخبز	ما أرديك إلا بسنة الله ورسوله صلى الله عليه وسلم وهما أردت
2	1001	التاجر وشدوا وثاقهم وطلعوا بهم إلى مدينة بغداد ، فعند ذلك جلس	وكانوا ثلاثمائة فارس مجتمعين من أوباش العربان فلما أسروهم أخذوا ما معهم
3	1001	ومتاع وأظهروا للتاجر أنهم من التجار وساروا معه فلما نزل أول منزل	به إلى بعض المدن لبيعه فيها فقتبعه رجلان من السكره وحملوا شيئا
4	1001	وعبيد وجوارا وخدم . ثم نفخ الخليفة المنديل وانبض الديوان فنزل علاه	وقال الخليفة لعلاء الدين : واره في التريا وخذ جميع ما تركه
5	1001	أو جاء أو غير ذلك وأنا خليفك في أهلك قال حيا وكرامة	موسى بن نصير في هذا الأمر ولك الراية البيضاء وكل ما تريده
6	1001	أبيه وغيره فافتقر حتى صار يشتغل مع الفعلة ، فمكث على ذلك	شيء رجع على بيع العبيد والجواري والإملاك وأنفق جميع ما كان عنده
7	1001	ومن نسب المال مكتسب والعز مرتجع إذا اشتكى المرء من داء ومن	أجابته هند بهذين البيتين : وما نبالي إذا أرواحنا سلسمت فما فقتناه
8	1001	أبي وكنت أدخرته لوكت الحاجة إليه فأنى كنت في زمن أبيك غنية	ولدي لا تظن أن هذا المال مال أبيك والله يا ولدي أنه

CQPweb v3.3.17 © 2008-2023

Help! for this screen

You are logged in as user [manis]

Figure 4: *Idgham* in CQPWEB.

4. Discussion

4.1. Derivation and Inflection in Arabic Morphology

Understanding the stem in Arabic is quite important

because this understanding will lead to the study of the word formation process in Arabic. Stem in Arabic is often equivalent to the term */jidzrul kalimah/* (جذر الكلمة) or */shighatul asās/* (صيغة الأساس) or often referred to as (base form), namely: *the part of a word to which an inflectional*

affix is or can be added^[2].

Regarding the study of derivation and inflection in Arabic, it should be noted that the difference between derivation and inflection in Arabic is not so clear, as in English and other languages. This is because Arabic morphology works on different principles, each language has its unique character. In addition, Arabic morphological theory shows the elements of word structure and sentence structure from different perspectives^[29].

On the other hand, Ryding also reminded us that without the basic principles of Arabic morphological theory, early learners of Arabic will have difficulty in utilizing and using Arabic dictionaries. Ryding also concluded that “Arabic morphology exhibits rigorous and elegant logic”. One of the uniqueness of Arabic morphology is that Arabic has a considerable number of discontinuous morphemes. Arabic morphology is fundamentally composed of a system of consonant roots that are connected to patterns of vowels in forming words or word stems. This system is not found in English. In more detail, the following is the opinion of Karin C. Ryding’s opinion regarding Arabic morphology.

It consists primarily of a system of consonant roots that interlock with patterns of vowels (and sometimes certain other consonants) to form words, or word stems. This type of operation is not unknown in English^[29].

The important elements in *Sharf* or Arabic Morphology are: (1) derivational morphology (such as the study of 10 verb forms in Arabic), (2) inflectional morphology (such as the paradigm of past tenses in Arabic), (3) *ishtiqaq* (etymology) which has more to do with the first element, namely derivational morphology. *Ishtiqaq* focuses on the study of word origins in Arabic and its development, and more specifically the study of *ishtiqaq* is centered on word formation (creation of words) derived from the lexical root system.

The important thing that becomes an element of distinction between Nahwu (syntax) and Sharaf (morphology) is that the study of morphology is not discussed case and mood. In the study of Arabic morphology, it is also discussed about “ablaut” or the procedure of distinguishing meaning with the influence of the word internal vowel (word internal vowel) or can also be studied in terms of “introflexion”, namely: a word internal change that sig-

nals a grammatical change. In English such as the words: man/men, foot/feet, sink/sank/sunk, and know/knew, these cases of word-internal vowel changes are quite common in Arabic, for example:

Katab-a (v) = ka + ta + ba (كَتَبَ) (he wrote)

Kutiba (v) = ku + ti + ba (كُتِبَ) (it was written)

Kutub (n) = ku + tu + bun (كُتُبَ) (books)

To validate stem words and inflected words, we can use The Quranic Arabic Corpus. In the corpus, there are several menus that can be selected by visitors or readers, such as: Quran Dictionary, which has branches such as: verb concordance, lemma frequency, and morphological search. This menu will be very useful for the study of word formation processes in Arabic, for example the results of searching for the word kataba in The Quranic Arabic Corpus:

The trilateral root kāf tā bā (ك ت ب) occurs 319 times in the Quran, in seven derived forms:

49 times as the form I verb kataba (كَتَبَ)

once as the form III verb kātibu (كَاتِبُ)

once as the form VIII verb ik'tataba (اُكْتُتِبَ)

260 times as the noun kitāb (كِتَابَ)

once as the noun kitābī (كِتَابِي)

six times as the active participle kātib (كَاتِب)

once as the passive participle maktūb (مَكْتُوب)

To strengthen the data, the analysis of the word *kataba* was added from the Leipzig corpus, in which some data was added in the form of a word graph, namely words that have close usage with the word kataba (source: <https://corpora.uni-leipzig.de/>).

Based on **Figure 5** below, we can conclude that the word /كتب/ has close usage with the following words: الكتاب، مقالات، سلسلة كتب، بعنوان، عدة، أنه الكتب، رسالة. Another thing to note importantly in the Arabic word formation process is the ablaut. Ablaut is a vowel change to mark various grammatical functions^[1]. Ablaut in Arabic is often equivalent to the term /tanāwubush-shawā'itu/ (إبدال الصوائت) or /ibdālu shawā'itu/ (تناوب الصوائت):

الصوائت في جذع الكلمة لإحداث تغيير في معناها أو وظيفتها، مثلًا وهو على نوعين: تناوب نوعي، وتناوب كمي (ring, rang, rung) تغيير^[11].

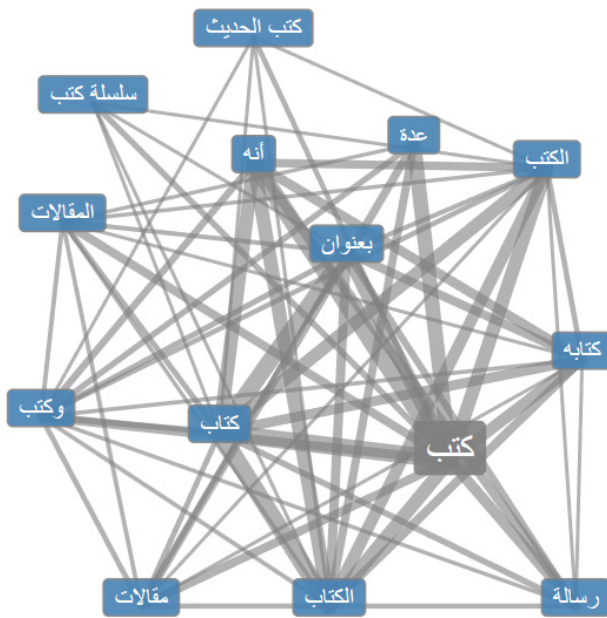


Figure 5. The Word Graph of “kataba” (write).

عدة الكتب (Several books)، بعنوان الكتب (Titled Books)،
(Hadith books)، سلسلة الكتب (Book Series)، الكتب (Books)،
، رسالة كتب (Message Books)، كتاب (Book) - كتب الحديث
، كتب المقالات (Article books).

From the definition of “ablaut” above, it can be concluded that ablaut is a vowel change (*taghyīrush-shawāit*) in the stem of a word that changes its meaning or function. There are also two kinds of ablauts, namely: qualitative ablaut and quantitative ablaut. In addition to the study of ablaut, Arabic morphology also studies words based on affixation, such as inflectional suffixes on words or stems, for example, *katab-at* (she wrote - she (woman) has writ-

ten) or example: kutub-an (books) as an example of the accusative case. In Arabic morphology, there is also the term ROOT, or three-consonant in the order (k + t + b) which is the ROOT of all word derivations meaning “to write”. This character is typical of the morphological system of Semitic languages. Ryding quotes that “the Semitic root is one of the great miracles of man’s language” ^[31]. Arabic is a Semitic language that has a rich morphology and relatively free word order. As such, Arabic can also be classified as a polysynthetic language ^[32].

In the case of ablaut in Arabic, we used the Tunisian Arabic corpus (<https://www.tunisiya.org/>). This corpus was chosen because it has quite a variety of contexts for the category of a word. Some of the word search categories offered in the Tunisian Arabic corpus are: blogs (323,339), category, conversation (phone), folktales (47,981), internet forum (96,355), jokes, literature (238,112), opinion (19,225), plays (25,079), poem, political speeches, proverbs, radio (13,774), recipes, religion, SMS/Facebook (25,203), songs, sprots, television (drama) (192,906), and web (77,138). To see the ablaut mapping of the words *ka-taba* and *katabat*, and *katabna*, see **Table 1**.

From the ablaut mapping results in **Table 1**, we can see that the ablaut shift process from *kataba* - to *katabat* and *katabna* has decreased in number or frequency of occurrence in the Tunisian Arabic corpus. It can be seen that the ablaut phenomenon causes a decrease in the frequency of occurrence in the text. The high frequency occurrences of the word *kataba* are in the categories of blogs, dramas, and internet forums. In these three categories, the word *kataba* has a fairly high frequency of occurrence.

Table 1. Mapping Ablaut Forms in the Tunisian Arabic Corpus.

[illegible]

4.2. Compounding

Compounding is the joining of two or more language forms so that they can have their function; for example, the joining of two morphemes, words, phrases, or clauses to form one construction. Compound is also usually defined as the combination of two existing/old words to form a new word, the ability to create new words by combining old words, such as the word blackbird which is composed of the words black (adj) and bird (noun) ^[33].

One of the most important word formation processes in Arabic, especially in terms of lexical types, is compounding into one word, which in Arabic, is often equivalent to the term *naḥt* (نحت), namely: composing a word by conjoining other words ^[29]. In some references, the term compounding as part of morphology is equivalent to the term *shiyāghah al-murakkabāt* (صياغة المركبات) ^[11].

In addition, the term compound itself is specialized in several categories, such as compound noun, compound adverb, compound adjective, compound personal pronoun, compound object, and compound subject. There are several subprocesses and variations of the compounding procedure in Arabic, although this procedure is not so common in the tradition of Arabic morphological data. This compounding procedure is widely used in Modern Standard Arabic (MSA: Modern Standard Arabic), especially in the process of forming new terms, and the process of loan translations, especially related to technical terms, uses a lot of compounding procedures.

Examples of this procedure in classical Arabic are often found in the following examples.

- such as the words /māwardi/ مَاورِد and /rismāl/ رِسْمَال which are derived from the words ماء وُرْد /mā'un/ (water) + /wardun/ (flower) = /māwardi/ (flower water)

- and رأس مال comes from the word /ra'sun/ (head) + /māl/ (treasure) = /rismāl/ (capital).

- Another example in classical language /ḥabaqur/ حَبَقُور (cold) which comes from the words /ḥabba/ + /qur-run/ = قُرْ + حَبْ = /ḥabaqur/ ^[34].

- Another example: lā (not) + (*markaziyya*) (centralization) = *laa-markaziyya* (decentralization) ^[29].

In corpus studies, compounding forms are rare, for example the word (مَاورِد), which is not found in the Leipzig corpus or the CQPWEB corpus. The word (māwardi) is found in the Tunisian Arabic corpus (**Figure 6**). This is evidence that words in compounding form are rarely used by Arabic speakers.

Some other examples also exist that show that there is an abbreviation in the compounding process, as in the word “supersonic” /faw-sawtiyy/ (فَوْصُوتِي), derived from the word /fauqa/ (فوق) (above) + /shautun/ (صوت) (sound) = (فَوْصُوتِي) (supersonic). The abbreviation process of the word /fauqa/ becomes /fau/ then combines with the noun /shautun/ and adds the suffix /iy/ as an adjective/nisbah marker in Arabic.

In Indonesian, the compound form can be composed of two words, such as (rumah) (house) + (susun) (flat) to (rusun) (flat house or In Indonesian, the compound form can be composed of two words, such as (rumah) + (susun) to (rusun). The following is an example of the search results for the word rusun in the Indonesian corpus (**Figure 7**).



Figure 6. Compounding Forms in Arabic Found in the Tunisian Arabic Corpus.

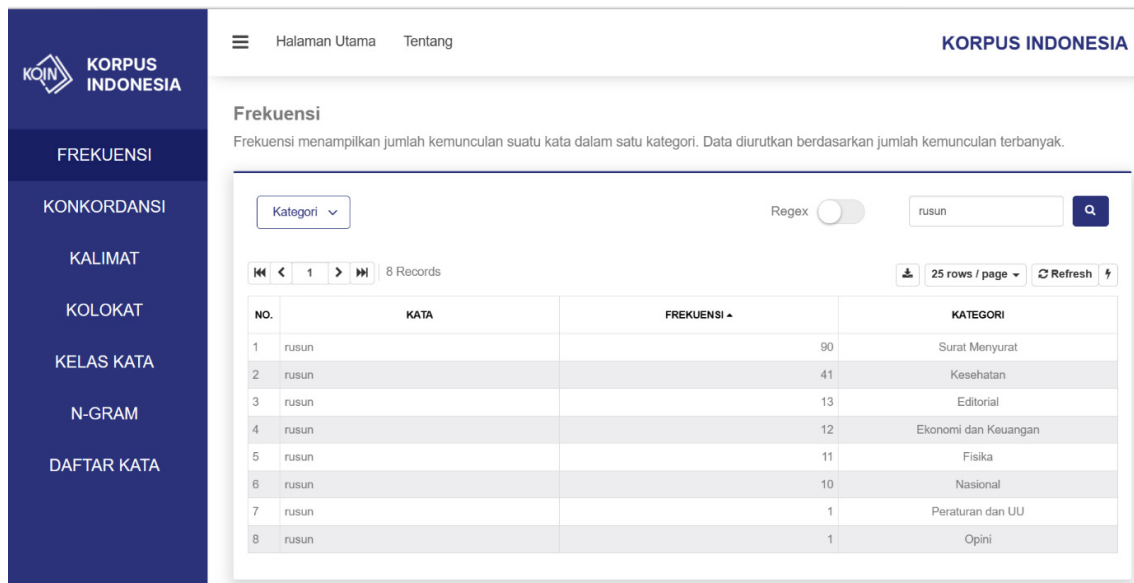


Figure 7. Examples of Compound Forms in Indonesian.

Source: korpusindonesia.kemdikbud.go.id.

Word: Rusun Number of Occurrences: 4,811

(Source: corpora.uni-leipzig.de).

From **Figure 8**, we can see that the word “rusun” (flats) has quite intense usage with several words such as: “perumahan” housing, “warga” residents, “pembangunan” development, jokowi (president of Indonesia).

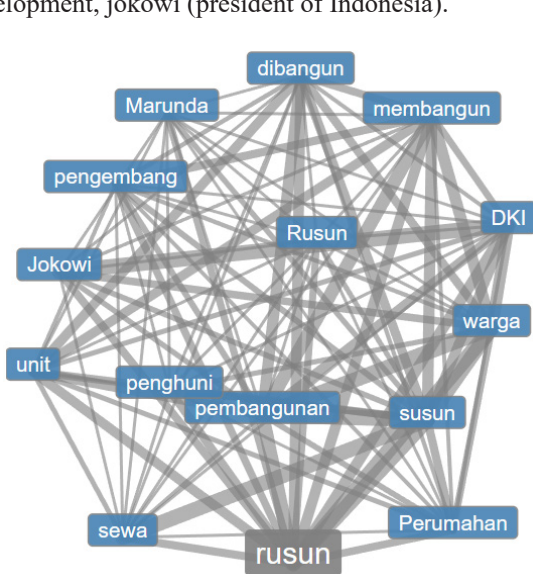


Figure 8. Examples of Compound Forms in Indonesian Word Graph.

Rusun (flats), rusun dibangun (flats built), Marunda, DKI (special capital region), pengembang rusun (flat developers), Jokowi, warga rusun (flat residents), unit rusun (flat units), penghuni rusun (flat residents), pembangunan rusun (flat builders), rumah susun (flats), sewa rumah susun (flat rental), perumahan susun (residential flats).

Based on the analysis above, it can be concluded that compound forms in Arabic and Indonesian are rarely

found. Compound in Arabic is dominantly formed from two words, while compound in Indonesian can be formed up to four words (**Figure 9**).

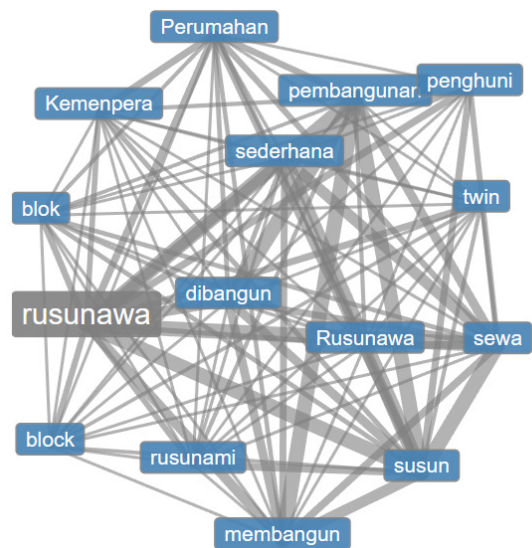


Figure 9. Compound Form in Indonesian Consisting of Four Words (Rusunawa) (Rumah Susun Sederhana Sewa) (House + Flat + Simple + Rental) in Word Graph.

Kemenpera (Kementerian Perumahan Rakyat) (ministry of public housing), pembangunan rusun (flat construction), penghuni rusun (flat residents), rumah susun sederhana sewa (simple rental flats), rumah susun sederhana milik (simple owned flats), blok, twin, membangun rusun (building flats).

Word: Rusunawa Number of Occurrences: 3,843.

(Source: corpora.uni-leipzig.de).

4.3. Abbreviations and Acronyms

The abbreviation is a morphological process in the form of dating one or more parts of a lexeme or combination of lexemes so that a new form that has the status of a word occurs. Abbreviations include abbreviations, beheadings, acronyms, contractions, letter symbols, and abbreviations^[1]. Some other examples of comparing words that are purely compounded and compounded forms that are abbreviated are as follows.

Amphibious = adj. which lives/travels in water and on land = /barmā'iyy/ (برمائي)^[35].

The word /barrun/ (برّ) (land) + /mā'un/ (ماء) (water) = /barmā'iyy/ (برمائي) amphibious. In the word /barmā'iyy/ the compounding element is dominant and no abbreviation process is found.

Prehistoric = the word /qabla/ (قبل) (before) + /tārīkh/ (تاريخ) (history) = qabtārīkhiyy (قبتاريخي). The word /qabla/ is abbreviated to /qab/ then combined with the word /tārīkh/ and added with the suffix /iyy/ as a nisbah marker.

In Arabic, compounding into two words (tarkīb) (تركيب) is also known, where the lexical item formed is not only one word in Arabic, but the new word formation becomes a noun phrase, as in the following example:

Multiracial = muta'addidul-ajnās (متعدد الأعناس) = whose members come from various races.

Multistorey = muta'addiduth-thawābiq (متعدد الطوابق) = with many storeys^[35].

In Arabic word formation, loanwords are also found, namely: several words (primarily nouns) in MSA that are borrowed directly from other languages, they cannot be broken down into roots and patterns. In this case, loanwords are positioned as words borrowed from other languages, and the word cannot be broken down into ROOTS and PATTERNS or word formation patterns in Arabic, as in the following words:

Chloroquine = كلوروكين^[36].

Some loanword cases in Arabic are rare, the word (Chloroquine = كلوروكين) is not found in CQPWEB, but is found in the word graph of the Leipzig corpus (Figure 10).

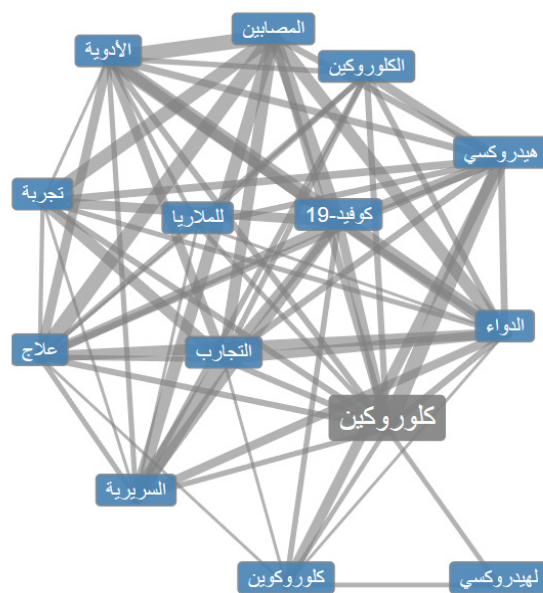


Figure 10. Examples of Compound Forms in Indonesian Word Graph.

،(Malaria) الملاريا ،(An experiment) تجربة ،(Infected people)

كلوروكين (Chloroquine)، الأدوية (Medicines)، المصابين

،(Clinical) كلوروكين (Chloroquine)، لهيدروكسي (For hydroxy)

الدواء (Medication)، هيدروكسي (Hydroxy)، السريرية

كوفيد ١٩ (COVID-19)

Word: كلوروكين Number of occurrences: 52

(Source: corpora.uni-leipzig.de)

Covid 19 = كوفيد 19^[36]

Coronavirus = فيروس كورونا^[36]

Hydroxychloroquine = هيدروكلوروكين^[36]

Loan words in the above examples are often referred to as ad hoc transcriptions, by applying basic orthographic principles from the source language to the loan words of the target language. Loan words in Arabic usually also show regional variation that reflects the local phonological situation, in the case of lexical variation it is usually different from the original, as in the following word examples:

(أوتيل) 'otel = hotel, the word ('otel) is a form of loan word that adjusts to the local phonological process. As the sound /g/ is considered hard so it becomes /j/, as in the word "garage" which is absorbed into (جرار)^[37].

Another word formation can also be seen in the phenomenon of acronyms in Arabic. Acronym (acronym) is a shortening process that combines letters or syllables or other parts that are written and pronounced as a word that meets the phonotactic rules of a language, such as FKIP (Fakultas Keguruan dan Ilmu Pendidikan) "Faculty of Teacher Education and Science" and ABRI (Angkatan Ber-

senjata Republik Indonesia) “Armed Forces of the Republic of Indonesia”^[1]. Studies related to acronyms in Arabic have been carried out previously^[38]. Some references explain acronyms in Arabic into one unit with the study of abbreviations, here are some examples of shortening that are often found in Arabic.

- اهـ (from the Arabic word /intahā/ (انتهى) “it has ended”) to mark the quotation.

- صلعم (from the full form /shallā Allahu ‘alaihi wa sallam/ (صلى الله عليه وسلم) “may God bless him and give him peace”).

Some examples of secular abbreviations in Arabic:

- Km (كم) kilometre = kīlū = (كيلو) = kilogramme
- M (م) for mitr (متر) metre
- Sh (ص) for a shafhah (صفحة) or “page”
- l (ل) for līra (ليرة) lira, sometimes also used /l.l/ (ل.ل) for līra lubnāniyya (Lebanese pounds)
- j (ج) for junayh (جنيه) pounds or guineas
- s (س) for santīm (سنتيم) centimetre

In addition to acronyms, in Arabic, it is also necessary to review some studies regarding clipping in Arabic. A comparative study of clipping in Arabic and English has been done before^[39]. Clipping is a shortening process that retains one part of the lexeme, such as Prof (professor), Mrs (mother), Bu (ibu), Mr (father), Pak (bapak)^[1].

Clipping is one part of the process of new word formation. Unlike English, the clipping process in Arabic tends to be a “partition process”. The study of clipping is also included in the realm of “word manufacture”, namely the process of word formation without using morphological, phonological, and orthographic motifs. The first time in the tradition of Arabic grammar, the study of clipping was discovered by Sibawaih, when Sibawaih first heard several Arabic speakers utter the following sentence.

لا تا فا بلى which has the full form
لا تفعل فا بلى افع
(Do not do it, yes you can do it)

Some Arabic language experts define clipping as: “idiomatic structure”, namely: the omission of a letter or more from the source form. Some examples of clipping in Arabic can be found in the following patterns.

Clipping in Arabic vocative: Like when calling Haris by name but only by Har, as in the following example: (يا

يا حارث) becomes (يا حار). Calling “hi Manshur” undergoes clipping to “hi mans”, (يا منصور) becomes (يا منص). Clipping in Arabic vocatives also occurs in words ending in ta’ marbutah (ة). As in the word (يا عائشة) only referred to as (يا عائش), or as in the word /musa-firah/ (مسافر ة) which means female traveler, only referred to as (يا مسافر). Clipping in Arabic can also be utilized in a stanza of poetry as in the following stanza, which is to create an aesthetic form of the final sound in the poem. *حبالكمو رماما ... و اوضحت منك شائعة اماما*. In the poem, the word imama (اماما) is originally imamah (امامة), to match the sound of the previous final sound, ramama (رماما), the ta’ marbutah (ة) in the word imamah is clipped.

Clipping also occurs in Arabic compound nouns, such as in the form of the compound noun: hadramaut, which is composed of = (حضر) + (موت), becomes a vocative by adding a vocative particle (يا) to become (يا حضر موت) which means “O Hadramaut”, then undergoes clipping to become (يا حضر) “O Hadra.”.

In a recent study, Raof developed the concept of translation in “jargon”^[40].

Jargon is a term used to denote language used in a particular field, where one cannot understand the term unless one knows the context or has studied in a related field. Jargon Translation is the translation of any foreign terms that do not exist in the Source Language (SL) where the terms refer to certain fields, such as science, professional terms, etc. This morphological process of jargon can be seen from the word formation process in English as follows:

The derivation is the formation of words by giving affixes or suffixes, e.g., un-happy

Acronyms, acronyms, or abbreviations of a term, e.g.: NATO, FBI, UN, etc.,

Backformation, the removal of a word affix, e.g., donation becomes donate.

Blending, combining the beginning of one word with the end of another word, e.g., motorcycle + hotel = motel.

Borrowing, borrowing a word from a foreign language, e.g., yogurt (Turkish).

Calque, borrowing by adjusting the culture, e.g., skyscraper becomes “skyscraper”.

Clipping is, an abbreviation of terms, for example:

Laboratorium abbreviated as Lab.

Coinage is a trade product name that can be used as a new word, e.g., nylon, jeans.

Compounding, combining two terms into one, for example, text + book = textbook.

Conversion is the conversion of word function if the grammar changes, for example, the word report is a verb, but can be used as a verb in the sentence “she wrote a good report”.

5. Conclusions

Arabic as one of the Semitic language families has a distinctive character in word formation. This can be seen in the process of word formation through five important stages, namely: 1) inflection, (2) derivation, (3) allomorphy, (4) abbreviations - acronyms, and (5) compounding. These five things play an important role in the process of word formation in Arabic. The study of Arabic morphology is certainly different from the study of morphology in other languages, but it cannot be denied that the theoretical studies (key concepts) in the development of the theoretical study of Arabic morphology have a lot of influence from the study of Semitic languages. It is based on this study that comparative analysis between Arabic and other languages can be done optimally.

Inflectional morphology explains how words have various forms, and how words undergo inflection to express grammatical categories such as singular/plural, or past/present tense. On the other hand, when “derivation” is positioned as the process of forming words or lexical units, then procedurally, derivation acts before the process of inflection. In this case, the derivation process occurs at the word stem level and modifies it.

The important elements in Sharf or Arabic Morphology are: (1) derivational morphology (such as the study of 10 verb forms in Arabic), (2) inflectional morphology (such as the paradigm of past tenses in Arabic), (3) ishtiqaaq (etymology) which has more to do with the first element, namely derivational morphology.

An important element in the process of Arabic word formation is allomorphs. In this study, allomorphs will intersect in the study of “assimilation” in Arabic. Sound assimilation occurs when several sounds become one and are pronounced as a double sound.

Word formation in Arabic is also through the process of compounding and acronyms. There are several subprocesses and variations of the compounding procedure in Arabic, although it is not very common in the tradition of Arabic morphological data. This compounding procedure is widely used in Modern Standard Arabic (MSA: Modern Standard Arabic), especially in the process of forming new terms, and the loan-translation process, especially related to technical terms, uses a lot of compounding procedures. Some references explain acronyms in Arabic as an integral part of the study of abbreviations.

After a comparative analysis based on corpus linguistics and morphology, it can be concluded that, some words in the form of stems, after experiencing the inflection process, the stem-shaped words have decreased the number of frequencies in the corpora. The limitation in this research is the use of some Arabic Corpora as follows: korpusindonesia.kemdikbud.go.id, the Quranic Arabic Corpus, Corpora Collection Leipzig University, Tunisian Arabic Corpus), and parallel corpora (Leeds Corpus of Contemporary Arabic: powered by CQPweb, The Qur'an (Arabic): powered by CQPweb, Arabic – English Parallel Corpus. Arabic corpora will continue to be updated, so that in future studies related to word formation processes, other Arabic corpora can be used. In general, a potential strength of future research is that the study of corpus engines will continue to grow, especially in relation to “artificial intelligence” studies. Therefore, corpus studies will have more potential in the future, because corpus studies have linked language studies with technology and artificial intelligence.

Funding

This work was supported by Non-State Budget Funds (APBN) from Universitas Sebelas Maret, for the fiscal year 2024, through the Applied Research Excellence scheme grant number [202/UN27.22/PT.01.03/2024].

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This research was funded by Non-State Budget Funds (APBN) from Universitas Sebelas Maret, for the fiscal year 2024, through the Applied Research Excellence scheme (PUT-UNS) with contract number 202/UN27.22/PT.01.03/2024.

Conflicts of Interest

The author declares no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

- [1] Kridalaksana, H., 2009. Dictionary of Linguistics Fourth Edition . PT Gramedia Pustaka Utama: Jakarta, Indonesia.
- [2] Richards, J.C., Schmidt, R., 2007. Longman Dictionary of Language Teaching and Applied Linguistics. Longman Group: London, UK.
- [3] Katamba, F., & Stonham, J. (2006). Morphology. PALGRAVE MACMILLAN.
- [4] Abdul Razak, Z.R., 2014. Word usage variations in Arabic newspapers: A corpus investigation. GEMA Online Journal of Language Studies. 14(2), 29–45. DOI: <https://doi.org/10.17576/GEMA-2014-1402-03>
- [5] Alromima, W., Moawad, I. F., Elgohary, R., & Aref, M. (2016). Ontology-based model for Arabic lexicons: An application of the Place Nouns in the Holy Quran. 2015 11th International Computer Engineering Conference: Today Information Society What's Next?, ICENCO 2015. <https://doi.org/10.1109/ICENCO.2015.7416338>
- [6] Alasmari, J., Watson, J. C. E., & Atwell, D. E. (2017). Using the Quranic Arabic Corpus for comparative analysis of the Arabic and English verb systems. International Journal on Islamic Applications in Computer Science And Technology, 5(3), 1–8.
- [7] Hadi, S., Chamamah, S., Ramlan, M., & Wijana, I. D. P. (2003). PHONOLOGICAL CHANGES OF ARABIC ABSORPTION WORDS IN INDONESIAN. Humaniora, XV(2), 121–132. <https://doi.org/https://doi.org/10.22146/jh.780>
- [8] Erwin, E., Loe, Y., 2017. Reduplication in Dengka Dialect of Rote Language: Generative Morphology Approach [in Indonesian]. Linguistik Indonesia. 17(1), 26–44.
- [9] Hakim, M.A., 2010. MORPHOLOGICAL STYLISTICS OF THE QURAN JUZ 30 . LINGUA: Jurnal Ilmu Bahasa dan Sastra. 5(1), 17–24. DOI: <https://doi.org/10.18860/ling.v5i1.610>
- [10] El-Affendi, M.A., 2018. The Generative Power of Arabic Morphology and Implications: A Case for Pattern Orientation in Arabic Corpus Annotation and a Proposed Pattern Ontology. Proceedings of the 5th International Symposium on Data Mining Applications; 21–22 March 2018; Riyadh, Saudi Arabia. Advances in Intelligent Systems and Computing. Springer: Cham, Switzerland. 753, pp. 36–45.
- [11] Baalbaki, R.M., 1990. Dictionary of Linguistic Terms English - Arabic. Dar el-Ilm Lil-Malayin: Beirut, Lebanon.
- [12] Khan, M.I., 2016. Reduplication in Arabic and Urdu. International Journal of English and Education. 5(4), 336–344.
- [13] Al Jarf, R., 2017. English and Arabic Word formation Processes for Translation Students [Course Material]. Available from: https://www.researchgate.net/publication/281003183_English_and_Arabic_Word-formation_Processes_for_Translation_Students (cited 25 May 2025).
- [14] Mubarak, F., Al-Khalifa, H., Al-Salman, A., et al., 2022. Arabic Phrases Errors in Writing Theses for Masters Students at Indonesian Universities. Eurasian Journal of Applied Linguistics. 8(3), 201–211. DOI: <http://dx.doi.org/10.32601/ejal.803016>
- [15] Joubran-Awadie, N., Shalhoub-Awwad, Y., 2023. Morphological distance between spoken Palestinian dialect and standard Arabic and its implications for reading acquisition. First Language. 43(2), 200–230. DOI: <https://doi.org/10.1177/01427237221145375>
- [16] Subramaniam, R., Kaur, S., 2021. A Corpus-driven Study on the Use of Passive Verb Bundles in Academic Writing: A Comparison between L1 and L2 English Speakers. GEMA Online Journal of Language Studies. 21(4), 64–87. DOI: <https://doi.org/10.17576/gema-2021-2104-04>
- [17] Sadjirin, R., Aziz, R.A., Nordin, N.M., et al., 2018. The development of malaysian corpus of financial english (MaCFE). GEMA Online Journal of Language Studies. 18(3), 73–100. DOI: <https://doi.org/10.17576/gema-2018-1803-05>
- [18] Rahim, H.A., Hasan, R.A., Hong, A.L., et al., 2021. The Diachronic Malaysian English Corpus (DMEC): Design, Development and Challenges. GEMA Online Journal of Language Studies. 21(4), 88–109. DOI: <https://doi.org/10.17576/gema-2021-2104-05>
- [19] Nor, N.F.M., Rahman, A.A.N.C.A., Jaluddin, A., et al., 2019. A corpus driven analysis of representations around the word ‘Ekonomi’ in Malaysian hansard

- corpus. GEMA Online Journal of Language Studies. 19(4), 66–95. DOI: <https://doi.org/10.17576/gema-2019-1904-04>
- [20] Hua, T.K., Khalis, H., Mohd-Said, N.E., et al., 2021. The polarity of war metaphors in sports news: A corpus-informed analysis. GEMA Online Journal of Language Studies. 21(2), 238–252. DOI: <https://doi.org/10.17576/gema-2021-2102-13>
- [21] Kulsitthiboon, S., Pongpairoj, N., 2018. Cooperative corpus consultation for acquisition of adjective + preposition collocations. GEMA Online Journal of Language Studies. 18(3), 57–72. DOI: <https://doi.org/10.17576/gema-2018-1803-04>
- [22] Hameed, A., 2022. A Corpus-Based Investigation of Saudi Corporate Discourse for Environmental Concerns. Eurasian Journal of Applied Linguistics. 8(2), 259–271. DOI: <http://dx.doi.org/10.32601/ejal.911559>
- [23] Moustafa, B.S.M., 2022. Thematisation Processes and the Representation of Love in The Forty Rules of Love: A Corpus Stylistic Analysis. Eurasian Journal of Applied Linguistics. 8(1), 298–312. DOI: <http://dx.doi.org/10.32601/ejal.911538>
- [24] Birhan, A.T., 2019. Review of Corpus Linguistics for Vocabulary: A Guide for Research. Eurasian Journal of Applied Linguistics. 5(3), 543–545. DOI: <https://doi.org/10.32601/ejal.651351>
- [25] Akinci, M., Yıldız, S., 2017. Effectiveness of Corpus Consultation in Teaching Verb+Noun Collocations to Advanced ELT Students. Eurasian Journal of Applied Linguistics. 3(1), 91–109. DOI: <https://doi.org/10.32601/ejal.461036>
- [26] Phoocharoenstil, S., 2021. Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. Eurasian Journal of Applied Linguistics. 7(1), 240–258. DOI: <https://doi.org/10.32601/ejal.911269>
- [27] Alkhuli, M.A., 2000. A CONTRASTIVE TRANSFORMATIONAL GRAMMAR: Arabic and English. Dar Al Falah: Amman, Jordan.
- [28] Kesuma, T.J., 2007. An Introduction to Language Research Methods. Penerbit Carasvatibooks: Yogyakarta, Indonesia.
- [29] Ryding, K.C., 2005. A Reference Grammar of Modern Standard Arabic. Cambridge University Press: Cambridge, UK.
- [30] Al-hamadi, H.M., Salman, I.M., 2011. Morphophonemic Interface in BBC English, and Basra Arabic: A Case Study of Phonological Conditioning. Journal of Basra researches for Human Sciences. 36(4), 81–109.
- [31] Lohmann, J., 1972. M. Heidegger's "ontological difference" and language. In: Kockelmans, J.J., (eds.). On Heidegger and Language. Northwestern University Press: Evanston, IL, USA.
- [32] Hussein, O., Rua, W., Sakha, T., 2023. Arabic as a Polysynthetic Language: Evidence From the Holy Qur'an. Theory and Practice in Language Studies. 13(3), 607–616. DOI: <https://doi.org/10.17507/tpls.1303.09>
- [33] Radford, A., Atkinson, M., Britain, D., et al., 2009. Linguistics An Introduction: Second Edition. Cambridge University Press: Cambridge, UK.
- [34] Adh-Dhāmin, 2007. Fiqhul-Lughah. Dārul-Āfāq al-'Arabiyyah: Cairo, Egypt.
- [35] Collin, P.H., Kassis, N., Angel, T., 1990. Harrap's English Dictionary for Speakers of Arabic. Kernerman Publishing, Inc: Montreal, Canada.
- [36] Organisation Arabe pour l'Education, la Culture et les Sciences, 2020. TERMINOLOGY DICTIONARY COVID-19 TERMS. Bureau de Coordination d'Arabisation: Rabat, Morocco.
- [37] Badawi, E., Carter, M.G., Gully, A., 2016. Modern Written Arabic A Comprehensive Grammar: Second Edition. Routledge Taylor & Francis Group: London, UK.
- [38] Rahman, A., Altakhaineh, M., 2017. The Linguistic Features of Acronyms in Arabic. Sage Open. October–December 2017, 1–14. DOI: <https://doi.org/10.1177/2158244017746710>
- [39] Oriabi, H.H., 2012. A Contrastive Study of clipping in English and Arabic. Journal of The College of Education for Girls for Humanities. 10. DOI: <https://doi.org/10.36327/ewjh.vi10.9151>
- [40] Raof, H.A., 2023. An Introduction to Arabic Translation: Translator Training and Translation Practice. Routledge: New York, NY, USA.