



**The morphological integration of loanwords  
into Modern Standard Arabic**

*Towards a morphological categorization of loanwords*

Niklas Gadeli

Bachelor thesis in General Linguistics

January 2015

Lund University

Centre for Languages and Literature

Supervisors: Gerd Carling, Andreas Hallberg

## **Abstract**

This thesis explores the morphological integration of Standard Average European (SAE) words into Modern Standard Arabic (MSA). The topic constitutes a challenge insofar as SAE and MSA are typologically very different, and integration of words from SAE into MSA should therefore be generally difficult. Loanwords were selected from a collection of contemporary short stories and filtered through a Modern Standard Arabic dictionary of words with non-Semitic origin. This was followed by a morphological analysis and categorization of the loanwords. The loanwords can be divided into two groups. Words in the first group do not fit into the so called root and pattern system of MSA, i.e. the configuration where a verbal root serves as the basis for derivations and inflections produced via internal vowel or consonant alternations. However, words in this group can take Arabic suffixes. The latter group can be subdivided into (i) loanwords which can be linked to formal roots as evident from their broken (= root internal) plural pattern, (ii) proper verbal roots or (iii) a combination of these. This classification enables us to create a scale of morphological integration. It is shown that suffixing (as in group 1) is not a viable strategy for integrating loanwords into a language exhibiting the root and pattern-system, such as MSA. Of special interest is the border between loanwords consisting of unintegrated solid stem words on one side, and on the other, loanwords that can be linked to a root.

## **Keywords**

Loanwords, borrowing, Modern Standard Arabic, morphology, broken plurals, root and pattern.

## **Acknowledgements**

I would like to express my gratitude to the people who supported me while writing this thesis. I am very thankful for the great involvement and encouragement of my supervisors Gerd Carling and Andreas Hallberg without whom this project would never have succeeded. I am also indebted to my family and friends who have motivated me to finalize the thesis in time.

## List of abbreviations

acc.	accusative
adj.	adjective
C	root consonant
CA	Classical Arabic
fem.	feminine
gen.	genitive
indef.	indefinite
m.	masculine
MSA	Modern Standard Arabic
n.	noun
nom.	nominative
pl.	plural
RL	recipient language
+rat	for rational beings
-rat	for non-rational entities
sg.	singular
SL	source language
v	(short) vowel
̄v	long vowel
v.	verb

## Table of contents

1. Introduction.....	5
2. Theory.....	7
2.1 Previous research.....	7
2.2 Loanword terminology and definitions.....	7
2.2.1 What is a loanword?.....	7
2.2.2 Reasons for borrowing.....	8
2.2.3 Different kinds of loans and grammatical borrowing categories.....	9
2.3 Some aspects of Modern Standard Arabic.....	10
2.3.1 The Arabic alphabet.....	10
2.3.2 Morphology: the root and pattern system in MSA.....	11
2.3.3 Morphology: the stem system of the triliteral verb.....	13
2.3.4 Morphology: number in MSA.....	15
2.4 The language academies.....	16
3. Method, material and procedure.....	18
3.1 Selection and cross-checking of the loanwords.....	18
3.2 Problems and solutions.....	19
4. Analysis of the corpus.....	21
4.1 General characteristics: elements common to all loanwords.....	21
4.2 Category 1: loanwords with solid stems.....	23
4.2.1 Word class.....	23
4.2.2 Solid stems.....	23
4.2.3 Spelling conventions.....	24
4.2.4 Number and gender.....	27
4.3 Category 2a: loanwords with abstracted roots.....	29
4.3.1 Plural pattern classes and root abstraction.....	29
4.4 Category 2b: loanwords with verbal roots.....	33
4.5 Category 2c: loanwords with a verbal root and a broken plural.....	34
4.6 Summary table: loanwords and the distribution of their MSA morphology.....	36
5. Discussion.....	38
6. Conclusion.....	41
7. Bibliography.....	45
Appendix.....	47

## 1. Introduction

Arabic is one of the largest languages in the world. It has around 300 million native speakers and is spoken as a second language (L2) by approximately 60 million people. The land area that is covered by this language is vast and stretches from Iraq and Khuzistan (in southwest Iran) in the East, to Morocco and the northeastern Nigeria in the West (Owens 2013: 2). Furthermore Arabic is the language of the Quran, Islam's holy book, which makes it the religious and liturgical language of all Muslims worldwide. It is also one of the six official languages of the United Nations, enjoying the same status as English, French, Spanish, Russian and Chinese (Holes 2004: 1).

Arabic is a language that is subject to the phenomenon of *diglossia*: one could say that there are two different forms of Arabic. On one hand we have the spoken dialects which constitute the mother tongues of all native speakers, and on the other hand Modern Standard Arabic (MSA) which is the language used in writing and formal speech, considered to be the modern descendant of Classical Arabic (Holes 2004: 3-6).

As mentioned by Pereltsvaig (2012: 92) Arabic belongs to the Semitic language family. More precisely both MSA, and the spoken varieties of Arabic, as well as Modern Hebrew and Samaritan all belong to the South Central branch, itself part of the Central branch of the Semitic languages.

Arabic is spoken in different dialects in different regions and some dialects have more in common than others. The dialects are less and less mutually intelligible the further apart we move geographically. One of the underlying reasons for this existence of a continuum of Arabic spoken dialects is language contact. This contact occurred between Arabic and the different languages spoken in the Middle East before the arrival of the Arabs, but also with neighbouring languages, and more recently, with languages of colonial powers (Pereltsvaig 2012: 94).

An important result of language contact is borrowing. The average borrowing rate of the *Loanword Typology Project* which aimed to study lexical borrowing patterns from a representative sample of languages all over the world, was of 24,2% (Tadmor 2009: 55). This means that nearly a quarter of all lexemes in the sample were loanwords, which is substantial (even though there is a bias in the sample towards languages that borrow more, as the authors admit).

This thesis will focus on a specific aspect of the borrowing process: the integration of loanwords into the morphology of Modern Standard Arabic. The main questions that we will try to answer are the

following: how are loanwords integrated morphologically into MSA? Why are some loanwords more easily integrated than others? This entails studying how elements from languages without any root and pattern system, such as SAE, are integrated in a language where this property is a design feature, namely MSA.

An observation made by myself is that the frequent presence of long vowels in the Arabic script could be an easy way to tell that a word is an unintegrated loan. This means that the actual spelling of a loanword is a very important clue as to its status: if it is written with many long vowels in MSA this would be a sign of its non-integration into the Arabic root and pattern system, whereas when a word contains few long vowels, this would be indicative of integration. This hypothesis will be tested throughout this paper.

The study will thus focus entirely on MSA, and elements of the Arabic spoken varieties will be given minimal attention. The material selected for the study consists of a collection of 12 short stories: *Modern Arabic short stories: a bilingual reader* (Newman and Husni, red. 2008) that I have read and carefully analyzed. The list of loanwords can be found in the appendix.

As the thesis should be readable for people unfamiliar with the Arabic script the words that are written in Arabic are also transcribed. There are several transcription systems that are used by arabists, but I have opted for the one used in the *Encyclopedia of Arabic Language and Linguistics* (Versteegh and Eid, red. 2009) for its clarity and popularity in the field of Arabic linguistics. It represents both the spelling and pronunciation of the Arabic script/alphabet (see Reichmuth 2009: 515-520 for further information about the transcription system). Some sections of the thesis might seem to have an Arabist bias, but I felt that a thorough description of Arabic language structure was needed, since this language is highly important for linguistic theory but not generally well-known to the Western reader. I believe that it would have been difficult to skip this step.

Borrowing in MSA is an interesting topic, and a challenge, because it has not been studied much. There is little, if any, genetic affiliation between the languages involved and their typology differs a lot as well, something which should have interesting consequences when it comes to integration of loanwords. The loanwords of this corpus are then the meeting point of SAE agglutinating or analytic structures with root and pattern (i.e. non-concatenative) Arabic language morphology. My aim is to study how words from the source language are incorporated morphologically into the recipient language given all the differences that exist between them.

## 2. Theory

### 2.1 Previous research

Previous studies in the area of loanwords in Arabic include Sa'id (1967), Smeaton (1973), Heath (1989) and Rolland (2014). However, their topic is not exactly the same and they differ in the extent to which they incorporate MSA in their research.

Sa'id (1967) entirely focuses on MSA and studies the grammatical treatment of loanwords with regard to gender, number, case, pattern congruity as well as loan derivatives. He covers many of the aspects that will be addressed in my thesis. Smeaton (1973) concentrates on the impact of technical terms from American English on the colloquial Arabic of the region of Al Hasa in Saudi Arabia and the lexical innovations that arose due to the contact situation. The study takes place in a context of intense contact between Hasawis and Americans working together in the Arabian American Oil Company (Aramco) but covers some loanwords from Persian, Hindustani, MSA and the Hijazi colloquial as well. More importantly, a part of it is dedicated to grammar relating to loanwords that affects both MSA and Hasawi.

Heath (1989) investigates the borrowing from French to Moroccan Colloquial Arabic (MCA) and from Classical Arabic to MCA. He gives very few references to borrowing into MSA but the study is nonetheless interesting as Heath and I happen to have some common entries in our corpuses.

Rolland's work (2014) is of a more lexicographical character, he compiled a dictionary of words in MSA that lack a Semitic origin by assembling information from several other dictionaries. His research proved to be particularly useful for me when assessing whether or not the words selected in my corpus were of foreign origin, i.e. whether they were loanwords.

### 2.2 Loanword terminology and definitions

#### 2.2.1 *What is a loanword?*

Language contact often leads to a change in the "structural inventory of at least one of the languages involved, and sometimes of both" (Matras 2009: 146), in other words a structure or form from one language is imported into another language. This is called "borrowing".

The metaphor of borrowing has been criticized as it is a bit misleading: a monolingual speaker of the community might not know that the word has a foreign origin, whereas borrowing suggests that it is a

fully conscious action. Other objections are that it will obviously not be returned as is the case of most borrowed items. The word *borrowing* also emphasises the ownership aspect more than the actual sharing of structures or forms (Matras 2009: 146). Even the word *transfer* suggests that the donor language loses an element (Haspelmath 2009: 37). But again, the term *borrowing* is used very often and most people understand what is meant by using it. It might be more ambiguous in theory than in practice and I will use this terminology throughout the thesis. *Borrowing* can also be called *copying* or *replication*. Synonyms to *borrowings* are *loanwords* or *loans*. The languages involved are referred to by the terms of *donor/source language* and *recipient language*, respectively.

Borrowing can actually correspond to two different situations, depending on whether the borrower is a native or a non-native speaker of the recipient language. In the first case, that is, if a native speaker uses structures from a foreign language and implements them in her/his mother tongue speech this is often called “adoption”. In the second case, that is, if a non-native speaker uses structures from her/his mother tongue in the speech of the foreign language this is referred to as “imposition” (Haspelmath 2009: 36).

Some researchers tend to define borrowing more restrictively only by what we just called “adoption”. “Borrowing” then denotes linguistic material that is transferred from a source language to a recipient language “via the agency of speakers for whom the latter is the linguistically dominant language, in other words, via RL agentivity”, RL standing for Recipient Language. This is contrasted with imposition, which is due to SL (Source Language) agentivity (Winford 2010: 171).

A warning must be issued: loanwords and native words live in a complementary relationship, we can only determine that a word is a loanword based on the knowledge that we have about a certain language. For native words the situation is ambiguous, it can never be ruled out that a native word was not in fact the result of a borrowing at some point of the history of the language, that we ignore. Haspelmath (2009: 38) says “Thus, we can identify loanwords, but we cannot identify ‘non-loanwords’ in an absolute sense. A ‘non-loanword’ is simply a word for which we have no knowledge that it was borrowed.” However, from the synchronic perspective that I am applying in this paper, the borderline between native words and loanwords is quite clear.

### 2.2.2 Reasons for borrowing

Matras (2009: 150) discerns between two reasons often mentioned for resorting to borrowing: so-called *gaps* in the structural inventory of the recipient language and the *prestige* of the donor language. Gaps in one language means that bilinguals or semi-bilinguals feel that there are things they



cannot express entirely in the language they are speaking and replicate a structure from the other. They permit themselves to use structures normally restrained to a certain language and a certain context in another language and another context. These gap-fillers are often *cultural loans*: loans that are brought to a speaker by new social activities and cultural acquisitions e.g. community functions and institutions, agricultural or food products and technical innovations. Cultural loans thus have an enriching effect on the lexicon of a language.

Borrowing motivated by prestige has a different mechanism: speakers seeking for approval and social status use elements of the speech of the socially more powerful and dominant community. They try to reproduce the associations that these words have in a new setting which would reflect positively on them. We can note that in contrast to the gap-fillers which do not possess any parallel expressions in the recipient language, most prestige loans have parallel expressions in the recipient language, which however lack the special effect looked for (Matras 2009: 150).

Finally, it is interesting to note that the social hierarchy does not have to be in favour of the donor language even if this is often the case i.e. the donor language can also be a small, less powerful language sometimes corresponding to a marginalized group or a minority group (Matras 2009: 151). An intra-linguistic example of this could be when well-educated people consciously use "bad language" because of its "covert prestige", as first noted by the American sociolinguist William Labov.

### 2.2.3 Different kinds of loans and grammatical borrowing categories

A distinction is often made between two basic ways of borrowing elements from one language to another: *matter* borrowing (MAT) and *pattern* borrowing (PAT) (Matras and Sakel 2007). This is sometimes also called *material* borrowing vs. *structural* borrowing (Haspelmath 2009: 38-39). Sakel (2007: 15) describes matter borrowing as a replication of "morphological material and its phonological shape" from one language in another language whereas in pattern borrowing "only the patterns of the other language are replicated" in other words the organization, distribution and mapping of grammatical or semantic meaning can be borrowed but the form itself is not.

Haspelmath (2009: 39) mentions other kinds of borrowings such as *calques* (loan translations), which are a common form of pattern borrowing consisting of an item-by-item translation of a complex source unit (often a compound) like the French *presque'île* from Latin *paen-insula* 'almost island'. Calques can also involve morphological derivatives or phrasal expressions. Another type is *loan meaning extensions* which consist of a polysemic pattern that is being replicated in the recipient language. *Loanblends* are hybrid borrowings containing both borrowed and native material.

The borrowing of grammatical elements follows a certain hierarchy, entirely based on frequency, and providing the likelihood for a certain word class to be affected by borrowing. Cf. the following list (from Matras and Sakel 2007: 61):

nouns, conjunctions > verbs > discourse markers > adjectives > interjections > adverbs > other particles, adpositions > numerals > pronouns > derivational affixes > inflectional affixes

This means that nouns and conjunctions are the elements that are most subject to borrowing whereas inflectional affixes are the elements that are least borrowed. That nouns occupy such a strong position can be explained by their *referential* function, they “cover the most differentiated domain for labelling concepts, objects, and roles” (Matras 2009: 172), and nouns are most prone “to express new concepts and to name objects and institutions” (Matras and Sakel 2007: 65). Nouns do dominate the loanwords found in the corpus (see 3.1 and 4.2.1).

## 2.3 Some aspects of Modern Standard Arabic

### 2.3.1 *The Arabic alphabet*

We shall here go through some basic aspects of the Arabic alphabet that will be useful for the reader and for the sections to come. As previously mentioned, the detailed information on Arabic language structure in the following sections may seem a bit out of place in a work on general linguistics, but I have deemed it necessary to include these facts in order to be able to carry out my subsequent argumentation. The information is taken from the first chapter of Haywood and Nahmad (1965 [2009]: 1-21) if not stated otherwise. First, we should note that Arabic is written from right to left. The script is cursive and cannot be written in any other way, in contrast to English where a printed script exists along with cursive handwriting. There is no distinction between uppercase and lowercase letters, i.e. no capitalization (Ryding 2005: 10). The alphabet consists of 28 letters (29 if the symbol called *hamza* is counted as a separate one) which are all consonants, but three of them are used as long vowels or diphthongs as well namely ا *ʾalif*, و *wāw*, and ي *yāʾ*.

Letters have different forms depending on their position. A letter is either written in isolated, initial, medial or final form. The letters ا *ʾalif*, د *dāl*, ذ *dāl*, ر *rāʾ*, ز *zāy*, و *wāw* are never joined to a succeeding letter, only to a preceding one. Visually there is a small blank space left after writing one of these letters. There are 6 vowels in MSA, 3 short ones (a, u, i) and 3 long ones (ā, ū, ī). We also find two diphthongs in MSA: اَو *aw* and اَي *ay*. The short vowels are diacritics written under and above the

consonants:  $\overset{\text{ˆ}}{\text{ـ}}$  *damma* pronounced *u*;  $\overset{\text{ˆ}}{\text{ـ}}$  *fatha* pronounced *a* and  $\overset{\text{ˆ}}{\text{ـ}}$  *kasra* pronounced *i*. The long vowels have been mentioned earlier, they are  $\text{ا}$  *ʾalif* ā,  $\text{و}$  *wāw* ū, and  $\text{ي}$  *yāʾ* ī.

As stated in Ryding (2005: 30-31), in practice short vowels are not written in text, with two exceptions: the Quran and children's schoolbooks. Short vowels are explicit in the Quran so that readers and reciters can be absolutely sure about the right pronunciation of the sacred text. They also appear in schoolbooks so that students can study and master word structure and spelling when they learn to read MSA. When reading skills improve, the use of short vowels in texts diminish and finally disappear. As the short vowel patterns are predictable, they are seen as redundant. When it comes to learners of Arabic as a foreign language, the absence of vowels is often a problematic phenomenon. Extra attention has to be given to word structure, morphological patterning, and memorization of the exact sound of a word and its spelling. Even if the vowels are invisible, they have to be pronounced.

This shows that for native speakers, vocalization patterns are most of the time obvious and intuitive. For non-native speakers and for an unknown word, theoretically every consonant could be followed by one of the three short vowels available. As it is practice to omit the short vowels, a clash can be expected when it comes to words for which adult native speakers ignore the vocalization patterns or know them less well. Loanwords are a typical category of such words, which we will discuss more in the following sections.

### 2.3.2 Morphology: the root and pattern system in MSA

According to Holes (2004: 99) the root and pattern is a crucial principle of Arabic derivational morphology. The root can be seen, for a majority of the words in the language, as a semantic abstraction consisting of three consonants,  $C_1C_2C_3$ , from which words are derived by the superimposition of templatic patterns. The words that are formed are both structurally and semantically related to the root.

To illustrate this we will use the often cited verb  $\text{كتب}$  *kataba* 'he wrote' or 'to write'. The 3rd masculine singular perfect form is the default form that would correspond to the notion of infinitive in English (along with the "verbal noun" or *maṣḍar*). This is also the common citation form for a root.

Below follow some other words (see Wehr and Cowan [1979] 1994) (henceforth "Wehr"): 951-952 for the full list) that have the root  $\text{كتب}$  k-t-b, all connected with the basic meaning of 'writing':

كِتَاب *kitāb* 'book'

كِتَابَةٌ *kitāba* 'writing'

كَاتِب *kātib* 'writer', 'writing'

مَكْتَب *maktab* 'office'

مَكْتَبَةٌ *maktaba* 'library'

مَكْتُوب *maktūb* 'letter', 'written'

We can note that all these words contain the root consonants ك ت ب k-t-b and have something to do with the basic idea of writing.

The roots need to interact with a pattern in order to create a stem and a word. The radical consonants represent the semantic 'root' and are set in a 'pattern' of vowels and affixes which expresses the grammatical category of the word (Badawi, Carter, and Gully 2004: 26). The grammatical category can be for example noun/agent or verb/imperfect/derived stem. The result of the root and pattern interaction is a stem to which additional affixes for number, gender, case/mood and definition are attached (Badawi, Carter, and Gully 2004: 26). Ryding clarifies the nature of the vowels and affixes of the pattern: "Therefore, the components of MSA pattern-formation include: six vowels (three long: /aa/, /iī/, /uū/; three short: /a/, /i/, and /u/); seven consonants ( , t, m, n, s, y, w); and the process of gemination" (Ryding 2005: 48). According to Badawi, Carter, and Gully (2004: 26) all Arabic words except most particles can be expressed as a root that is set in a pattern. They observe that the number of radical consonants is most often three, but sometimes four and rarely two (it can be five and six but only in non-Arabic roots something which is relevant for the present work).

The patterns of the words above are then the following: (C denoting the consonantal roots,  $\bar{v}$  a long vowel):

كَاتِب *kātib* 'writer' has the pattern of an active participle in stem I, CāCiC.

مَكْتَب *maktab* 'office' and مَكْتَبَةٌ *maktaba* 'library' have the patterns of the noun of place maCCaC(a).

كِتَاب *kitāb* 'book' and كِتَابَةٌ *kitāba* 'writing' are two nominalizing patterns (the latter also a verbal noun pattern) of the form CiCāC(a).

مَكْتُوب *maktūb* 'letter' has the pattern maCCūC which is the passive participle pattern for stem I, and could also mean 'written' (See also *Noun patterns* in Badawi, Carter, and Gully 2004: 49-50, 88-90).

The root concept is essential for understanding the structure of an Arabic dictionary or lexicon. The Arabic dictionary is traditionally and most often not based on spelling but rather on roots. Each entry starts with a root and beneath it follow all the words derived from that same root. But the roots in the dictionary are organized alphabetically i.e. the roots follow the order of the letters of the alphabet (Ryding 2005: 49). This means that to use an Arabic dictionary correctly it is necessary to have some knowledge of the root system in order to identify the roots in the word that we want to look up but also to know the order of the alphabet to orientate oneself between the different root entries and find the actual root.

The organization of words that share the same root in the same entry applies to words that are genuinely Arabic or very arabicized. Loanwords that show no morphological integration are found alphabetically in the lexicon, since they do not possess any root under which they can figure (Ryding 2005: 49). We can thus see a dichotomy in the organization of the Arabic lexicon between words that on one hand are Arabic or treated as Arabic belonging to root entries, and on the other hand words that are treated as foreign and unintegrated, which appear alphabetically and do not belong to any root entry. It is the border between those two categories of words and the content of the said categories that I aim to study further.

### *2.3.3 Morphology: the stem system of the trilateral verb*

An example of the information conveyed by short vowels is the difference between active participle and passive participle in the stem II-X that is differentiated morphologically only by the vowel associated with the second radical: *i* for the active participle and *a* for the passive participle. So for the verb in stem II عَلَّمَ *allama* ‘to teach’ or ‘he taught’ the active participle is مُعَلِّمٌ *mu‘allim* meaning ‘teaching’ or ‘teacher’ and the passive participle is مُعَلَّمٌ *mu‘allam* ‘taught’.

The term *stem* in Arabic derivational verb morphology does not correspond to the more common usage of this word. The derived verb forms “extend or modify the meaning of the root form of the verb, giving many exact shades of meaning” (Haywood and Nahmad 2009: 151). Stems are represented by Roman numerals and there are arguably nine different patterns that constitute the stem system in use, namely forms I-X (form IX being uncommon). The root form of the verb, most often represented by the three root consonants, constitutes stem I. An important thing to notice is that “each derived form is associated with certain meaning patterns” (Haywood and Nahmad 2009: 151). Below, we will list the forms I-X in perfect with the verb in stem I كَتَبَ *kataba* ‘to write’ or ‘he wrote’, accompanied by the meaning patterns and some examples (adapted from Haywood and Nahmad

2009: 152-153). The root consonants are in red in the transcription, the pattern of infixes and vocalization is in black. When a root consonant is doubled, the first consonant is marked in red.

Table 1

Form N°	Perfect	Meaning pattern	Examples
I	كَتَبَ <i>kataba</i>	Basic form of the verb. Irregular vocalization pattern.	كَتَبَ <i>kataba</i> ‘to write’
II	كَتَّبَ <i>kattaba</i>	Strengthening/intensifying. Causative; transitive of intransive roots.	I عَلِمَ <i>‘alima</i> ‘to know’ becomes II عَلَّمَ <i>‘allama</i> ‘to teach’
III	كَاتَبَ <i>kātaba</i>	Performing the action to another person. Attempting the act.	III كَاتَبَ <i>kātaba</i> ‘to write to’
IV	أَكْتَبَ <i>aktaba</i>	Transitive of intransitive v. Causative of transitive v. “Stative verbs” derived from nouns.	IV أَعْلَمَ <i>‘a‘lama</i> ‘to inform’
V	تَكَتَبَ <i>takataba</i>	Reflexive of II (sometimes I). Verbs derived from nouns of quality/status. To represent oneself having the quality of the root meaning.	I كَسَرَ <i>kasara</i> ‘to break’, V تَكَسَّرَ <i>takassara</i> ‘to be broken’
VI	تَكَاتَبَ <i>takātaba</i>	Reflexive of III, often implying mutual application of the action.	VI تَكَاتَبَ <i>takātaba</i> ‘to write to one another’
VII	اِنْكَبَ <i>inkataba</i>	Passive sense (perhaps originally reflexive).	VIII اِنكَسَرَ <i>inkasara</i> ‘to break’ (intransitive)
VIII	اِكْتَبَ <i>iktataba</i>	Reflexive of I, for varied twists of meaning from the root idea.	I نَفَعَ <i>nafa‘a</i> ‘to profit, benefit (trans.) VIII اِنْتَفَعَ <i>intafa‘a bi</i> ‘to profit by’.
X	اِسْتَكَبَ <i>istaktaba</i>	Believing someone or something has the quality of the root. Asking for the quality or act of the root.	I حَسُنَ <i>hasuna</i> ‘to be good’, X اِسْتَحْسَنَ <i>istahsana</i> ‘to think good, admire’

### 2.3.4 Morphology: number in MSA

Pluralization of nouns is achieved in MSA by two different means: either by suffixation or by so called interdigitation of consonant and/or vowel patterns on the root consonants of the singular noun, thus following a certain pattern and generating what is called the “broken plural” or “ablaut plural” (Holes 2004:162). The dual and collective forms exist as well but will not be treated here.

The MSA suffixes for the two genders available in the language are different. For the masculine the suffixes are  $\text{ونَ} -\bar{u}na$  in nominative and  $\text{ينَ} -\bar{i}na$  in accusative and genitive. For the feminine the suffixes are  $\text{اتُ} -\bar{a}tun$  in nominative and  $\text{اتٍ} -\bar{a}tin$  in accusative and genitive. While the masculine plural suffixes are added directly to the noun stem, the feminine plural suffixes are added after removal of the feminine sg. suffix  $\text{ة} -a$  (Holes 2004: 163).

The broken plural works differently, basically it consists of different morphological processes (i.e. a pattern) applied to the roots of a singular noun in the same ways as described earlier i.e. by vowel insertion, consonant gemination or consonant affixation to the consonantal roots (Holes 2004: 168). Broken plurals are thus roughly equivalent to internal regular sound alternations as illustrated by singular *foot* vs. plural *feet* in English. This plural inflection is central to Arabic morphology and it is more or less predictable depending on the nature of the noun, some can be predicted from the shape of the singular noun whilst others cannot.

Let us illustrate this by the earlier examples that possess a broken plural form:

$\text{كِتَابٌ} kitāb$  ‘book’ -> pl.  $\text{كُتُبٌ} kutub$  ‘books’. The pluralization pattern is CvCāC (sg.)-> CuCuC (pl.).

$\text{كَاتِبٌ} kātib$  ‘writer’ -> pl.  $\text{كُتَّابٌ} kuttāb$  ‘writers’, ‘writings’. The pluralization pattern is CāCiC (sg.) -> CuCCāC (pl.).

$\text{مَكْتَبٌ} maktab$  ‘office’ -> pl.  $\text{مَكَاتِبٌ} makātib$  ‘offices’. The pluralization pattern is CvCCvC (sg.) -> CaCāCiC (pl.) (*m* is not a root consonant though, but this word uses a 4 radical pattern).

In section 4, we will see what happens when loanwords meet the requirements of MSA morphology such as described above. But before that, I would like to present some sociolinguistic facts about the Arabic language.

## 2.4 The language academies

There are several language academies in the Arabic speaking world, the oldest one is the Arab Academy of Damascus which had its first meeting in 1919. More recently similar language academies have appeared in Cairo, Baghdad, Amman and Rabat (Holes 2004: 309). They all have three basic functions: (1) to protect the Arabic language against dialectal influence (2) to clear the language of foreign lexical elements brought in via press, radio or writers and (3) to adapt the language to the modern needs, particularly in the fields of science and technology. Arabic is viewed as a “pure” language that has to be protected against harmful influences. It is believed that the dialects have been corrupted by contact with foreign languages and that, at some remote point in history, they were entirely pure (Holes 2004: 310). However, Arabic has never been “pure” nor really homogeneous. In the earliest texts, including the Quran we can note a morphological and syntactical variability, surely due to different geographical origins of the Quran scribes (western Arabia versus eastern Arabia) and there are undeniable cases of foreign lexical borrowing as well. Foreign vocabulary continually found its way into Arabic, with a very strong import under the Abbasid power during the 9th and 10th centuries for instance (Holes 2004: 310). Holes points out that the “purity” of literary Arabic is a myth, from a literal historical point of view, because the language has always been changing. The reactions of the academies appeared in response to an intensified pace of change, but the change has been following the Arabic language since the beginning.

Different strategies have been used in order to adapt Arabic to the needs of the modern world. In many cases the coinage of new terms that would correspond to the multitude of foreign elements and the delicate levels of difference between them was too difficult a task for the academies. The new words were to be derived from the root and pattern system or by extending the semantic range of archaisms (Holes 2004: 311). Some proposals met with little success. Archaisms such as *ʾirzīz* ‘sound of rain, tremor’ for *telephone*, replacing *tilifūn*, and *jammāz* ‘swift-footed camel’ for *tram* were proposed by the Cairo Academy but never spread. They did not work as they did not correspond to the active set of roots in the language (the roots were used infrequently) nor were they semantically transparent for non-academicians (Holes 2004: 311).

Other coinages were more successful, the ready-made patterns with different semantic meanings such as the nouns of instruments *miCCaC(a)* and *miCCāC* have been used for coining words such as *miṣʿad* ‘elevator’ (from *ṣaʿada* ‘to climb, go up’) or *mijhar* ‘microscope’ (from *jahara* ‘to become perceptible, visually or acoustically’) (Holes 2004: 311-312), often thanks to greater transparency.



Building a standardized and internally coherent technical and scientific vocabulary is another challenge, due to the the few Arabic affixes corresponding to the Greco-Latin “in-”, “para-”, “hypo-” etc. which is solved by paraphrasing and sometimes leading to very heavy constructions. An example mentioned by Holes (2004: 312) is the word “indivisibility” which along with countless other similar constructions needs to be arranged in such a way that the different morphemes of the source words become long noun phrases. Below follows the example adapted to our transcription system and with the words in MSA added by myself:

Source language: indivisibility

Recipient language: عدم القابلية لتجزئة *‘adam alqābiliya litajzi’ā*

‘lack (of) the-susceptibility to the-divison’

### 3. Method, material and procedure

#### 3.1 Selection and cross-checking of the loanwords

The first step consisted in finding a good source from which I could extract loanwords.

There were several factors that influenced the choice: because of the format and nature of the bachelor thesis it seemed most appropriate to study written text (instead of spoken MSA in political speeches for instance). I then had to make sure that the text was entirely in MSA and void of influences from the spoken varieties of Arabic that would affect the results. In order to process a larger amount of text it seemed preferable to have a bilingual text i.e. a source in MSA with a translation in English, so less time would be spent on translating and more on analyzing.

*Modern Arabic short stories: a bilingual reader* (Newman and Husni, red. 2008) was chosen based on the criteria above. It contains 12 short stories written between 1973 and 2004 by authors from all over the Middle East. It has an English translation following the stories in Arabic, a brief author biography, a discussion of the stories' context and background as well as a very useful glossary containing, amongst others, words coming from the vernaculars and to be avoided for this study. The corpus is then both representative and extensive, as well as based on idiomatic material.

Based on the definitions of loanwords of Matras (2009), Matras and Sakel (2007), Haspelmath (2009) and Winford (2010) (see 2.2.1 and 2.2.3) I created a list of words taken from the short stories that purportedly qualified as loanwords. These words were then checked against the dictionary in *Les mots de l'arabe moderne d'origine non sémitique* (referred to hereafter as "Rolland") by the French linguist Jean Claude Rolland (2014), published as a part of his PhD thesis and, to my knowledge, the only existing modern etymological dictionary that has been written by a scholar on MSA. That dictionary enabled me to make sure that the words previously selected were of non-Semitic origin. However, it is not a complete dictionary. Indeed, examples, dates and history regarding the different words as found in a normal etymological dictionary are sometimes lacking while the emphasis lies on the etymons and source languages.

All potential loanwords were then cross-checked with a French etymological dictionary (<http://www.cnrtl.fr>) and an English etymological dictionary (<http://www.oed.com>), above all in order to account for the words of foreign origin that were not part of Rolland's dictionary. This also enabled me to confirm that the words that were part of Rolland were of foreign origin. After deleting some words that did not qualify as loanwords, I had a list of actual loanwords in MSA.

In order to study how well the loanwords were integrated in the Arabic root system I used Wehr as a tool for analysis. Words in *A dictionary of modern written Arabic (Arabic-English)* (Wehr and Cowan 1994) that belonged to the entry of either a verbal root, or had a broken plural form were treated as if they had been given a Semitic root. Words that had neither verbal root affiliation nor a broken plural and were written in the alphabetical order in the dictionary (otherwise organized after verbal roots) were considered not to have been given a Semitic root. The list was further refined as to evaluate why such a difference existed between loanwords possessing a root and loanwords lacking this property. Another important tool, used for a detailed analysis of the broken plurals, was Ratcliffe's classification system of broken plurals (see 4.3.1).

Some might argue that loanwords used in a construction specific to the RL cannot count as loanwords any more. As Haspelmath (2009: 37) says, "...when a word is analyzable within the recipient language, it can normally not be a loanword, because it was created within the recipient language". This principle was problematic for the current study and I chose not to apply it here, as this is exactly the process that we want to study. In order to respect that principle we would have to eliminate a lot of loanwords that could be analyzed from a root and pattern perspective, thus invalidating the aim of the study and its feasibility. Apart from this important exception, the general definitions of 2.2.1 and 2.2.3 have been followed throughout the thesis.

Finally, I should mention that the detection of loanwords in the short stories was facilitated by the fact that the SL very often was either French or English, two languages that I know well. When a certain Arabic word resembled the word in the English translation, it was also possible to directly infer that it was a loan, as SAE and Arabic are not genetically affiliated to each other, and most of the loans are from SAE languages. It became clear after a while that the borrowings were mostly nouns and sometimes adjectives (in total 51 nouns and 6 adjectives), so I could focus mainly on those elements while reading. In general, the loanwords were often quite transparent and easy to recognize.

### 3.2 Problems and solutions

Arabic lexicology has unfortunately a long way to go. The ideal situation would have been to run all the words from the short stories through an etymological database. As the latter does not exist yet I had to follow the steps mentioned above, along with some intuition and knowledge of MSA to select the loanwords in the first place. For practical reasons I focused on *matter* borrowings (Matras and Sakel 2007) which are the only ones that can be detected. Having double checked the words with Rolland's dictionary, a French etymological dictionary and an English etymological dictionary I can

be sure that all the words provided in the final list are actual loanwords. However, it is of course possible that I did not recognize all the loanwords in the short stories. The aim with the word list in this thesis is not to create a loanword database with statistical ambitions but rather to illustrate and categorize the different means of integrating loanwords into the grammar of MSA.

In some cases where the words figured neither in Wehr nor in Rolland or where no information was provided on plural forms, I used the Arabic corpus search tool *ArabiCorpus* (hereafter “AC”). It is available online on <http://arabicorpus.byu.edu> and allows searching for the presence of a word in an MSA corpus consisting of newspapers, modern literature, nonfiction, Egyptian colloquial and premodern literature (or one of these categories). I used the categories *newspapers* and *modern literature*. The mere existence of a word in the corpus could justify that a loanword is used but might be too new for Wehr and Rolland whose works are based on older material. When it comes to plural forms, if a search resulted in very few tokens, say 2 or 3, I judged that it was not common to use a word in the form that I had put it in. On the contrary, if the search resulted in over a hundred (or even a thousand) of tokens, I judged it more probable that the word is used in that specific form in MSA contexts.

Objections can of course be raised when it comes to the choice of dictionary. *A dictionary of modern written Arabic (Arabic-English)* (Wehr and Cowan 1994) is the main dictionary used in this thesis for the analysis, in the absence of a better suited one. This dictionary is not ideal for several reasons, first of all it is quite old. The fourth edition was printed in 1979 and reprinted in 1994. It constitutes an “enlarged and improved version of the original corpus” (Wehr and Cowan 1994: v), first published in English in 1961 and taking into account the earlier German versions *Arabisches Wörterbuch für die Schriftsprache der Gegenwart* from 1952 and a supplement from 1959. Some of the words of the corpus are too recent to be part of the dictionary. Second, the arrangement of the entries is not entirely homogeneous as new entries and revisions sometimes have been added directly without consideration to the format of the older entries, “no attempt was made to follow through consistently with all possible changes in presentation and arrangement” (Wehr and Cowan 1994: v). However, the dictionary is the most comprehensive Arabic-English dictionary available and has been widely praised for its scholarship, accuracy and reliability. It is an indispensable reference work and tool for studying MSA.

#### 4. Analysis of the corpus

I compiled a list of loanwords which included the following information: the word written in MSA in the form taken from the short story, its transcription, translation (in the short story), different grammatical features and some other information (see appendix). I also created a summary table of the loanwords and their MSA morphology (see 4.6, table 4).

A few elements were common to all loanwords, irrespective of the level of integration into the Arabic root and pattern system. After a first analysis of the loanwords it was also clear that some characteristics were shared only by some of the loanwords. This enabled me to conduct a categorization of the loanwords into 4 different groups based on the level of integration into the Arabic root and pattern system. We will look more closely into what unites and distinguishes different words from each group. The list of the loanwords in the appendix should be consulted when reading this section and the following ones for a better understanding.

The level of integration was measured by evaluating how central the MSA morphology of loanwords was to MSA grammar. Loanwords without a root and a pattern were judged to be least integrated and loanwords with a root and a pattern were judged to be most integrated. Loanwords that had a broken plural pattern but extracted roots without any meaning (purely formal roots) were judged less integrated than loanwords possessing a productive verbal root. This resulted in a hierarchy of loanword integration into MSA.

The four categories with increasing level of integration and their main morphological traits are:

1. Loanwords with solid stems, plurals in *-āt* and several long vowels (no roots)
- 2a. Loanwords with broken plurals, possible abstraction of 3 and 4 radicals (formal roots)
- 2b. Loanwords with verbal roots, mostly 4 radicals (“real” roots)
- 2c. Loanwords with verbal roots and broken plurals, mostly 4 radicals (“real” roots)

##### 4.1 General characteristics: elements common to all loanwords

A few characteristics are common to all of the loanwords of the list, and show that on a superficial level there has, in all cases, been an adaptation to the Arabic language. All of the loanwords are written in Arabic script and follow the Arabic rules of phonotactics and writing. This is not as trivial as it may seem since it is possible to write loanwords in Arabic using Latin script. Latin script provides a full vocalization of the word but presupposes some knowledge of the Latin alphabet on

behalf of the reader. But this has not been done for any of the loanwords of the corpus.

All loanwords are *adoptions* (Haspelmath 2009: 36) as the writers who have used the loanwords are all native Arabic speakers. The borrowings have then occurred via RL agentivity (Winford 2010: 171). Moreover, the loanwords entirely respect the syntax of MSA.

All loanwords obey the rules of definiteness, which is achieved by adding the definite article *al-* as a prefix, الباص *al-bāṣ* 'the bus' for instance. This may seem to be a weak sign of integration, but it should nevertheless be mentioned. Some prepositions are added as prefixes as well and if a definite article exists we get PREP-DET-N, بالهستيريا *bi-l-histiriyā*, (lit. *on-the-hysteria*) 'on hysteria' (*al* is pronounced *l* because of the preceding vowel sound). The same thing applies to the linking element و *wa* 'and', والصابون *wa-ṣ-ṣābūn* 'and the soap' (lit. *and-the-soap*) which is also written together with the succeeding word, as a prefix.

Loanwords can also be subject to suffixation, the personal pronouns are suffixed, abiding by the normal rules of MSA, شالها *šāl-ha* 'her scarf' (lit. 'scarf-her'). The derivational process of adjectivization in MSA can also be observed: the so-called *nisba adjectives* are created by addition of the suffixes -ي *-īy* (m. sg.) and -ية *-iyya* (fem. sg.) to existing nouns: البلاستيكي *al-blāstīkīy* 'plastic' (adj.).

When it comes to case, the situation is a bit more complex. There are three cases in MSA: nominative, accusative and genitive. Case is marked with suffixed short vowels at the end of words (Badawi, Carter, and Gully 2004: 33, 50). As mentioned earlier (cf. 2.3.1), short vowels do not normally appear in written text. This makes it impossible to study the use of the different cases as the case marking is often not visible in written MSA. There are some forms where it shows, e.g. the masculine suffixal plural endings and the dual endings or the accusative of words ending in a consonant. Unfortunately these forms were not used for any loanwords so they could not be studied either; generally there is not much that can be said about how the loanwords handle the case system.

Loanwords are used for constructions that are typical of MSA grammar such as the *idāfa* construction, a genitive construction, where two words are juxtaposed: the first term (the possessed) is in the so-called *construct state* without definite article, whereas the second one (the possessor) is most often definite. This often corresponds to compounds in the SL. In our corpus we find روب حمامها *rūb ḥamām-hā* '(her) bathroom robe' (lit. 'robe bathroom-her') which is a loanblend in the *idāfa*

construction. The second case of *idāfa* construction is أفلام البورنو *'aflām al-būrnū* 'porn movies' (lit. 'movies of the-porn') where actually both words are loans but unequally integrated from a morphological point of view. The postposition of adjectives after nouns is also respected for all the loanword adjectives of the corpus and is yet another example of MSA syntax being applied to the loanwords as if they were native elements.

Finally, words that have phonemes that do not exist in MSA have similar ones in the recipient language. There are three phonemes that do not exist in MSA but are often part of the source language's phoneme inventory: /p/, /g/ and /v/. The corresponding phonemes in MSA are /b/, /ɣ/ (spelled ġ in our transcription system) and /f/ as in the words كمبيوتر *kumbiyūtir* 'computer', غاز *ġāz* 'gas' and تلفزيون *tilifiziyyūn* 'television'. Sometimes /g/ is rendered by /ʒ/ as in سجارة *siġāra* 'cigarette'. This concludes the common characteristics of the loanwords.

## 4.2 Category 1: loanwords with solid stems

### 4.2.1 Word class

This category is the largest one, comprising 38 loanwords of a total of 57, thus representing almost 67% of the loanwords in the corpus. 32 of the 38 loanwords in this category are nouns and the remaining 6 are adjectives. This distribution is representative of the frequency of borrowing mentioned earlier (cf. 2.2.3) with nouns having a key position in the process of borrowing except for the fact that adjectives come before verbs and discourse particles, not the other way round. However, in MSA the aforementioned *nisba* adjectives are a set of adjectives that are derived from nouns by suffixation, and this class of adjectives is very closely related to the word class of nouns. In CA it was originally used for indicating membership of a class, tribe or location (Badawi, Carter, and Gully 2004: 55). This has been generalized in MSA, and nowadays any noun can easily be turned into an adjective by *nisba* suffixation.

### 4.2.2 Solid stems

All the words of category 1 display what in Arabic linguistic tradition is called *solid stems*, i.e. they are not analyzable in terms of root and pattern. Solid stems in Arabic consist of function words (prepositions, conjunctions etc.), pronouns and loanwords. They are listed in an Arabic dictionary according to their spelling (Ryding 2005: 50). Being solid stems these words are not productive, i.e. it is not possible to derive any words from them (except for *nisba* adjectives sometimes). This is in

contrast to words in Arabic that do possess a root and which are highly productive, to the extent that from a certain root we can get a multitude of lexemes only by changing pattern.

I have some speculations regarding these words, but they have not been proven in any way. As solid stem words can only appear in one specific form (or two if we count both singular and plural) they must be more difficult to remember for a native Arabic speaker. For words possessing radicals the multitude of patterns make sure that we will use a certain root many times when producing different words, and maybe keep the lexical knowledge of the root active in some way. The patterns are also a great help when it comes to memorizing words for L2 students of MSA. From one single root combination we can get a whole array of meanings, of words interconnected with each other. When a root is not productive, this has to cause a certain suspicion to the native speaker. If the words that they learn cannot be connected with other words in their mental lexicon, the solid stem word surely stands out as odd, exhibiting a foreign character. This might also mean that these words are harder to acquire in the first place.

#### 4.2.3 Spelling conventions

As we saw above, there are both short and long vowels in MSA, where short ones are often elided in writing. There seems to be some common agreement that loanwords should be spelled entirely with long vowels, “a general principle for notating foreign vowels whether long or short is to use a long vowel” (Badawi, Carter, and Gully 2004: 18) and it seems that this applies even more to words with solid stems. This is clearly visible in the corpus. The words below of category 1 were spelled with long vowels for every vowel sound slot in the MSA word (without counting affixes):

Table 2

أوبرا <i>ūbrā</i> 'opera'	جينز <i>jīnz</i> 'jeans'	كريمات <i>krīmāt</i> 'creams'
باص <i>bāṣ</i> 'bus'	ديكور <i>dīkūr</i> 'set' ('décor')	كولونيا <i>kālūniyā</i> 'cologne'
برونزية <i>brūnziyya</i> 'tanned'	روتين <i>rūtīn</i> 'routine'	كولونيل <i>kālūnīl</i> 'colonel'
بلاستيكي <i>blāstīkīy</i> 'plastic'	رولات <i>rūlāt</i> 'rolls'	موسيقى <i>mūstīqā</i> 'music' (n.)
بلوزات <i>blūzāt</i> 'blouses'	سيرك <i>sīrk</i> 'circus'	موسيقية <i>mūstīqiyya</i> 'music' (adj.)
بورنو <i>būrnū</i> 'porn'	غاز <i>gāz</i> 'gas'	موف <i>mūv</i> 'purple'
تيليب <i>tiyūlīb</i> 'tulips'	كاميرا <i>kāmīrā</i> 'camera'	نايلون <i>nāylūn</i> 'nylon'



There are some words that have multiple spellings. The following words were not spelled with only long vowels in my corpus but there were such variants in the dictionary:

بيرة *bīra* 'beer' can be spelled بيرا *bīrā*.

جغرافيا *juġrāfiyā* 'geography' can be spelled جيوغرافيا *jiyūġrāfiya*.

دش *duš* 'shower', دوش *dūš*.

هستيريا *histiryā* 'hysteria' either هستيريا *histīriyā* or هيستيريا *hīstīriyā*.

In the short stories, spelling seems to oscillate between versions with more or fewer long vowels. In some cases there does not seem to be any standard version of the noun but rather many different spellings that circulate. The different spellings are accounted for in the appendix, both when several spellings occurred in the short stories and when it occurred in the dictionary.

The words spelled entirely with long vowels represent 21 of the 38 words of the category in question, and if we add the 4 words above which also possess an alternative spelling with long vowels we get 25 words out of 38 that can be spelled with long vowels only, or 66% of the words of category 1. This does not look like a coincidence but rather like a tangible trend.

The other words of category 1, i.e. those that are not spelled with long vowels only have at least one short vowel sound and sometimes more. It could be that they are older loans or loanwords more frequently used, something which would maybe make their vocalization pattern more established among the Arabic-speaking population. This is difficult to prove though, especially without a complete etymological dictionary, and therefore this issue can unfortunately not be treated within the scope of the present thesis.

Generally speaking, category 1 thus seems to be characterized by loanwords written exclusively with long vowels. We should remember that long vowels are *always* written and short vowels are not. It seems that the long vowels are written out not so much in order to indicate vowel quantity or stress (although these factors might influence spelling as well) but rather to represent oral vowel quality. Given that loanwords are words that are less integrated generally in the vocabulary of a native Arabic speaker than native words, and that knowledge about them might vary from speaker to speaker, I believe that they are written out simply in order to tell the speaker which vowel sound to produce. In the same way as L2 students of MSA have a hard time knowing what short vowels are to be pronounced in a given word when reading an unvocalized text, exactly the same problem should

appear for native monolingual Arabic speakers if the loanwords were not spelled with long vowels.

Sa'id (1967: 42) recognizes the importance of spelling when it comes to loanwords:

The spelling factor has, in general, been ignored by students of linguistic interference, but it plays an important part in the transmission of loanforms into Arabic especially, since they are generally transmitted through the printed word. Among monolinguals adopting loanforms, the pronunciation of the loanform is based on imperfect spellings that leave out crucial information such as the indication of short vowels or the doubling of consonants.

Spelling loanwords with long vowels could hence be a conscious strategy of writers and publishers in order to avoid such confusion. It should however be pointed out that spelling characterized by long vowel sounds could complicate the task of identifying roots as, first of all, the longer the word, the more difficult to tell which consonants are the ones forming the root because of a greater choice of plausible options. Second, the vowel sounds *و* *wāw* and *ي* *yā'* count as both vowels and consonants (semi-vowels) which means that they might be root consonants or not, depending on their specific function in a certain word (*ا* *'alif* does not belong to this group). Short vowels being absent in written MSA text, we have no clue as to whether these sounds should be seen as root consonants or not. It is reasonable to assume that this double nature of the vowels is another factor that might complicate the task of identifying roots for the class 1 words.

Third, and most importantly, a profusion of long vowels may also inhibit the word in singular to fit in an MSA pattern, where, as we will see (cf. 4.3.1, table 3), the long vowels are mostly affixes of a specific pattern. Having too many of them might mean that it is impossible to insert the word into any pattern. The orthographic rules applied to loanwords then obstruct the integration of these words into the morphology of MSA. This theory, even though it is quite simple, makes a lot of sense. The spelling of loanwords in MSA takes on a whole new importance that it surely deserves.

The last and the least probable theory is that Arabic speakers who have seen the loanwords written in the source language have noticed that they are written with vowels and simply transfer these vowels into visible vowels in Arabic script, i.e. long ones, with the desire to create an equivalent that is close graphically. It might even be that they (consciously or unconsciously) transfer the vowel quantity system of Arabic to the source language and believe that the vowels that they see are long vowels, adopting an MSA view on the foreign lexemes.

#### 4.2.4 Number and gender

Smeaton (1973: 36) notes that the system of feminine plural *-āt* endings “functions as a general device for the pluralization of nouns of foreign origin which have not been assimilated into Arabic beyond the phonological stage.” This confirms once again the affiliation of the words in category 1 to a layer of superficial integration. The pluralization in *-āt* is also used by Smeaton as a factor of determination between different levels of integration. For nouns the switch from plurals in *-āt* to internal pluralization (i.e. broken plurals) is a clear sign of naturalization of the loanwords into the Arabic morphological system (Smeaton 1973: 61). The gender of the noun is not a criterion in Smeaton’s different degrees of loanword naturalization. Gender has generally not been seen as an important factor in previous studies with regards to morphological integration. However, this does not mean that it is not important.

From a morphological point of view, the masculine gender is the basic form in MSA whereas feminine gender is marked by suffixation (Ryding 2005: 119). Fem. sg. words often have a suffixal fem. pl. (and rarely broken plurals) whereas masculine words have either a suffixal m. pl. (only for masculine animates) or a broken plural (for animates and inanimates). However, when it comes to loanwords of category 1 the situation is more complex and the gender of the plural forms is maybe more a sign of integration than the gender of the singular form. Several words in category 1 are masculine in their singular form but feminine in their plural form. The feminine plural would then correspond to words that are seen as foreign or unintegrated irrespective of the gender of the singular noun.

The abundance of suffixal feminine plural endings in *-āt* of the loanwords in category 1 is striking. The nisba adjectives always have suffixal masculine and feminine endings in plural and thus have to be excluded as they do not have the two declination options of nouns into either suffixal plural forms or broken ones (there are some exceptions to this but they do not concern the words in our corpus). To illustrate this, the suffixal plural of the m. sg. برونزي *brūnzīy* ‘tanned’ is برونزيون *brūnzīyūn* and the suffixal plural of the fem. sg. برونزية *brūnzīyya* ‘tanned’ is برونزيات *brūnzīyāt*, but there are no broken plural forms that can be created.

This leaves us with 33 words, 18 of which are documented to have feminine suffixal plural forms, and there are in addition to these, a few words with an undocumented pluralization pattern, possibly adopting the suffixal feminine plural as well. It would thus seem that the feminine plural suffix is a kind of default plural marker which can be used for nouns which are masculine as well as feminine in

the singular. Its presence on loanwords must be a sign of non-integration in MSA, because otherwise at least some of the words in question would have displayed another type of plural.

Borrowed nouns ending in an *-ā* sound are often treated as feminine (Ryding 2005: 123). In category 1 some of these words already have a documented suffixal plural form but these ones do not: بطاطا *baṭāṭā* 'potato', جغرافيا *juġrāfiyā* 'geography', هستريا *histiriyā* 'hysteria', and سينما *sīnamā* 'cinema'. If these words are feminine in singular we could assume that they will have a suffixal feminine plural. For جينز *jīnz* 'jeans', it surely has the gender and number of the broader semantic category of *banṭalūnāt* 'trousers', in other words suffixal fem. plural. *Banṭalūnāt* also precedes the word *jeans* in the short story. دش *duš* is spelled دوش *dūš* in Wehr, and has a suffixal fem. pl. form. It is then probable that دش *duš* also has a suffixal fem. pl. form. بيرة *bīra* 'beer' having a feminine singular form should have a plural suffixal feminine form as well. To sum up we have 18 words out of 33 that have a documented suffixal feminine plural form and on top of that 7 words that are assumed to have a suffixal feminine plural. Regardless of how we count, the suffixal feminine plural seems to dominate the pluralization of the loanwords of category 1.

For some words we need to be cautious because the influence of semantics is considerable and a simple analysis of pluralization forms might not always be the best option. There are (a) words that are always borrowed and used in plural such as اليكترونات *ilīktrūnāt* 'electronics' because the concept in the SL is one of plurality, and that separates these words from the ones having a clear singular and plural form. (b) Collective nouns which are always pluralized by suffixal feminine endings in MSA (to refer to many single specimens in a group, such as "4 pieces of bread") and that then only have one possible plural form by definition such as maybe بطاطا *baṭāṭā* 'potato', غاز *ġāz* 'gas', موسيقى *mūsīqā* 'music' etc. (c) Nisba adjectives that mostly do not possess any broken plural forms at all and (d) words which have an unclear and maybe nonexistent plural form such as بورنو *būrnū* 'porn' and نايلون *nāylūn* 'nylon', probably for semantic reasons.

Loanword 1-6 (see 4.6, table 4) are words from category 1 where no suffixal fem. pl. form could be found and need to be commented. Except for *porn* and *nylon* just mentioned we have سيرك *sīrk* 'circus' which is probably not used much in plural generally; كئار *kanār* 'canary' (canary-bird) seems quite uncommon, and might therefore not have any clear plural neither. تيوليپ *tiyūlib* 'tulips' when translated takes a plural but has a seemingly singular form in MSA. It might be that it is rarely used in sg. so that the sg. form corresponds both to *a tulip* and *tulips*, like a collective noun. Finally, كولونيل *kūlūnīl*

*kūlūnīl* ‘colonel’ seems to lack any kind of plural, no suffixal fem. pl. nor suffixal m. pl. was detected in AC or Wehr even though it does not seem strange to talk about “colonels”. It is surely used more often in singular as well. This noun is the only animate of category 1 and is semantically masculine, which could explain why it is difficult to put it in a fem. pl. form. In this way, the gender of the loanword could maybe influence the pluralization pattern and the integration of the loanword. However the expected suffixal m. pl. form (for masculine animates) was not encountered either so it is difficult to draw any conclusions from this. The small number of animates in the corpus makes this phenomenon difficult to study as well.

Every time a clear suffixal feminine plural form exists, or an assumed one, this is annotated in the summary table (4.6, table 4) and in the word list in the appendix.

It is difficult to tell in what direction this goes. Does the fact that these words are considered to be solid stems engender suffixal feminine plural endings and a very clear spelling with long vowels because of an inability to comply with the root and pattern system inhibiting internal pluralization and automatic spelling of vowels? This hypothesis does not explain why they become solid stems. The other way round would be that these words are seen as so foreign that they are given a default plural form that is used for many different categories of nouns (see Ryding 2005: 134-140) and written with long vowels in order to signal that these words are somehow special, which would mean that their status as solid stems would be an epiphenomenon.

To sum up, category 1 is composed by loanwords having solid stems i.e. words that are unanalyzable from a root and pattern perspective. They are often spelled with long vowels and often have a feminine suffixal plural.

#### 4.3 Category 2a: loanwords with abstracted roots

##### *4.3.1 Plural pattern classes and root abstraction*

Common to category 2a, 2b and 2c is that they can all be linked to a consonantal root. The words in category 2a comprise 12 nouns. In contrast with the preceding category where the loanwords lacked any root affiliation, the loanwords of this one have come one step closer to an integration in the morphology of MSA. The loanwords of category 2a have one thing in common: they possess a broken plural form. This means that they are analyzable in terms of a root and a pattern. By studying the pattern that has been used to go from the singular form to the plural one, it is possible to abstract the

roots of the word. In the corpus these words are mostly masculine but there is one exception: سجارة *sijāra* ‘cigarette’ which is a feminine word. As I said before, broken plurals in the feminine are quite rare, the majority of the words that have a broken plural are masculine.

We will use the classification system of Ratcliffe (2008: 439-447 and earlier work) taking into account more than 90 percent of the different types of the very vast broken plural system. Broken plurals have been divided into 7 different classes based on their inflection pattern, the summary (see table 3) is presented as in Ratcliffe (2008: 444). The first column states the class of broken plural, then comes the word structure(s) in singular and finally the broken plural pattern(s). The table looks very technical at first sight but we will use the loanwords from the corpus to make it clearer.

To understand the table correctly we should clarify that C stands for root consonant, v for short vowel,  $\bar{v}$  for long vowel,  $\bar{u}$  for a long *u* and so on. The ending (at) represents the letter  $\text{š } \bar{t}\bar{a}$  ‘*marbūṭa*’.

Gemination is represented by two underlined root consonants. The forms in parentheses count for less than 10 percent of the class as a whole, the forms in square brackets are very rare (less than 10%).

Table 3

-----  
The Arabic ‘broken’ plural system

i.	CaCC	>>	CuCūC, ’aCCāC, CiCāC, [’aCCuC], (CīCān)
	CvCC	>>	’aCCāC, CuCūC, (CiCaCat)
	CvCvC	>>	’aCCāC
ii.	CvCCat	>>	CvCaC, CvC(a)Cāt
	CaCCat	>>	CaC(a)Cāt, CiCāC
iii.	CvCCvC(at)	>>	CaCāCiC
	CvCC $\bar{v}$ C(at)	>>	CaCāCīC, CaCāCiCat
iv.	C $\bar{v}$ CvCat	>>	CawāCiC
	CvC $\bar{v}$ Cat	>>	CaCā’iC
	C $\bar{v}$ C $\bar{v}$ C	>>	CawāCīC
v.	CāCiC (n.)(-rat.)	>>	CawāCiC
	CāCiC (n.)(+rat.)	>>	Cu <u>CC</u> āC, CaCaCat, (CuCāt)
	CāCiC (adj.)	>>	Cu <u>CC</u> aC
vi.	CvCāC	>>	’aCCiCat, CuCuC
	CaCūC	>>	CuCuC, ’aCCiCat
	CaCīC (n.) (-rat.)	>>	’aCCiCat, CuCuC
	CaCīC (n.) (+rat.)	>>	CuCaCā’, ’aCCiCā’
	CaCīC (adj.)	>>	CiCāC, [CaCCā]
vii.	’aCCaC	>>	CuCC, CuCCān

-----

In the appendix the words of 2a are grouped by alphabetical order in the same way as in the dictionary but below we will summarize the words of category 2a with regards to their different class affiliation which enables us to extract a root:

*Class i* comprises several inflection patterns. The ones encountered in the corpus are:

CvCC -> 'aCCāC: فلم *film* 'film' -> pl. أفلام *'aflām*, abstracted root: فلم f-l-m

فرن *furn* 'oven' -> pl. أفران *'afrān*, abstracted root: فرن f-r-n

Nouns with the pattern CṽC in sg. even though they are not shown in the table, are part of class i and ordinarily analyzed as having an underlying glide as second root consonant (Ratcliffe 2008: 441). For شال *šāl* it appears to be *ya* 'y' in Wehr. We have the two following examples in our corpus:

CūC -> 'aCCāC: روب *rūb* 'robe' -> pl. أرواب *'arwāb*, abstracted root: روب r-w-b

CāC -> CīCān: شال *šāl* 'scarf' -> pl. شيلان *šīlān*, abstracted root: شيل š-y-l

The broken plural patterns of class i thus give rise to a trilateral root.

*Class iii* has the following representatives in the corpus:

CvCCṽC(at) -> CaCāCīC: بركان *burkān* 'voclano' -> pl. براكين *barākīn*, abstracted root: بركن b-r-k-n

قنديل *qindīl* 'lamp' -> pl. قناديل *qanādīl*, abstracted root: قندل q-n-d-l

كرتون *kartūn* 'cardboard' -> pl. كراتين *karātīn*, abstracted root: كرتن k-r-t-n

CvCCṽC(at) -> CaCāCīCat: دكتور *duktūr* 'doctor' -> pl. دكاترة *dakātira*, abstracted root: دكتور d-k-t-r

And then we have a form which is a mix: دينار *dīnār* 'dinar' -> pl. دنانير *danānīr* indeed has a broken plural of class iii but a singular of form iv (sg. CṽCṽC -> pl. CawāCīC). The abstracted root could then be either دنر d-n-r or دنر d-n-r. Common for the loanwords of class iii is an abstraction of a quadrilateral root.

*Class iv* is manifested in two words:

CṽCṽC -> CawāCīC: كواليس *kūlīs* 'wing' -> pl. كواليس *kawālīs*, abstracted root: كلس k-l-s

*kawālīs* is the actual word of the corpus, translated as 'the wings' in the sense of 'backstage'.

CvCṽCat -> CaCā'īC: سجارة *sijāra* 'cigarette' -> pl. سجاير *sajā'ir*, abstracted root: سجر s-j-r.

This pattern then enables an abstraction of a trilateral root.

Class v takes the last element of category 2a:

CāCiC -> CawāCiC: كابل *kābil* ‘cable’ -> pl. كوابل *kawābil*. This word is written with a suffixal feminine plural ending in the corpus ( as كابلات *kābilāt*) but as it also had a broken plural I chose to insert it in category 2a. The class v pattern enables as well an analysis of the word as possessing a trilateral root.

There are actually 3 words from this category that possess both a broken plural and a suffixal fem. pl., loanwords 34-36 (see 4.6, table 4) كرتون *kartūn* ‘cardboard’, كابل *kābil* ‘cable’, and شال *šāl* ‘scarf’. It is difficult to explain this finding. It could depend on a lot of different factors such as personal preferences or regional differences. When loanwords from any of the subsections of category 2 have suffixal fem. pl. this is difficult to interpret, it could mean that they are less integrated but at the same time they do possess the characteristics that make them belong to one of the subsections of category 2 and this was judged more important for the sake of classification. A plausible explanation may well be that they are in the middle of the integration process and that they are moving from one category to another, here from category 1 to category 2a.

To sum up the root abstraction that can be made after studying the class affiliation of the different loanwords of the corpus we get:

- 4 words belonging to class i with an abstraction of 3 root consonants
- 4 (or 5) words belonging to class iii with an abstraction of 4 root consonants
- 2 (or 3) words belonging to class iv with an abstraction of 3 root consonants
- 1 word belonging to class v with an abstraction of 3 root consonants

This means that more than half of the loanwords with a broken plural pattern can be analyzed as having a trilateral root and less than half of them having a quadrilateral root. Generally in MSA trilateral roots are common and as seen earlier (cf 2.3.2) represent the basic structure of the MSA morphology system. Quadrilaterals are pretty common as well and they subdivide into words with Arabic origin and words from other languages (Ryding 2005: 93).

It is difficult to integrate words of more than four consonants in the root and pattern system but when those take a broken plural it is of the form CaCāCiC (Ratcliffe 2008: 442). This is achieved by deleting or ignoring one of the root consonants as in the examples from the same source:

CvCCvCvC -> pl. CaCāCiC: *barnāmiġ* ‘program’ -> pl. *barāmiġ*

*zanbarak* ‘[metal] spring’ -> pl. *zanābik*



Thus for all loanwords with four or more consonants that could possibly be analyzed as roots, class iii is the broken plural pattern used. The quadrilaterals therefore have a key position when it comes to loanwords and loanword integration and in two different ways: if the word possesses a broken plural form it often follows a quadrilateral pattern and if the word has been assigned a verbal root, it is often a quadrilateral one (see next section).

#### 4.4 Category 2b: loanwords with verbal roots

There are 5 words in the corpus that have a root entry in the dictionary, without having any explicit broken plural. We will here list the roots and derivatives.

The word *تليفون* *tilifūn* 'telephone' has the root *تلفن* *talfana* 'to telephone' i.e. t-l-f-n. A spelling variant that is recorded is *تلفوني* *tilifūnī*. The only other word with *talfana* as root in the dictionary is *تلفوني* *tilifūnī*, the nisba adjective meaning 'telephonic', it is maybe not very productive then, based on the words present in the dictionary. We also notice that *تليفون* *tilifūn* has a documented suffixal fem. plural which is a sign of non-integration, but here a real root can be identified, which on the other hand is a sign of integration.

The word *تلفزيون* *tilifiziyyūn* 'television' has the root *تلفز* *talfaza* 'to televise, transmit by television' i.e. t-l-f-z. Derived entries include the more arabicized version *تلفزة* *talfaza* 'television', the relative nisba adjective, and a noun with a similar meaning *تلفاز* *tilfāz* 'television set'. Actually, all the words above except *تلفزيون* *tilifiziyyūn* form an entry and the latter is the next entry in the dictionary along with its nisba adjective. However, I see no point in separating these two entries as all the words obviously can be connected to the root *تلفز* *talfaza* 'to televise, transmit by television'. The existence of two separate entries is surely a result of an error or the fact that entries are sometimes added on earlier entries in Wehr without rewriting the whole entry. The root seems productive, mostly when it comes to nouns.

*أسفالت* *asfalt* 'asphalt' has the root *سفلت* *saflata* 'to cover with asphalt, to asphalt' or s-f-l-t. This root is productive based on the words present in the same entry as we also find the verbal noun *سفلتة* *saflata* 'asphalt paving, asphaltting' and the past participle *مسفلت* *musaflat* 'paved with asphalt, asphalted' as part of the same entry.

The next word is *إسمنت* *ismant* 'concrete, cement' with the root *سمنت* *samnata* 'to cement' or s-m-n-t.

Other entries are سمنتو (Sp. cemento) ‘cement’ probably pronounced *samanto*, and the nisba adjective اسمنتتي *asmantī* ‘cement’ (adj.). The root seems minimally productive.

صابون *ṣābūn* ‘soap’ contrasts with the previous entry as it belongs to a root that is more productive, possessing more derived forms in the dictionary, alike the entries of fully Arabic words. The root is صبن *sabana*, s-b-n, which in stem II means ‘to soap, rub with soap’. Other lexemes with this root are صابونة *ṣābūna* ‘a cake of soap’ (nomen unitatis), صابوني *ṣābūnī* ‘soapy, soap-like, saponaceous, made of soap’, صبان *ṣabbān* ‘soap boiler’ and مصبنة *maṣbana* ‘soap works’. We can note that Rolland (2014: 220) gives صوين *ṣawbana*, ṣ-w-b-n as root with the meaning ‘savonner’ (‘to soap’), i.e. a quadrilateral root for the word صابون *ṣābūn* ‘soap’. This is an additional argument in favor of an existing root affiliation of the word; be it a trilateral or a quadrilateral one, the two different authors believe that the word has a root.

The degree of productivity as expressed earlier is based entirely on the number of lexemes that have been encountered as part of the root entry for the previous loanwords. Technically all loanwords possessing a verbal root could be equally productive and are able to generate all the different derivatives that could be expected from an Arabic noun with a verbal root.

So these five loanwords possess proper lexical roots, mostly quadrilateral, and can be distinguished from the previous categories by being productive. The verbal roots that have been identified in these loanwords have resulted in other words such as nisba adjectives, verbal nouns, participles, and other stems than I.

We can note that the meaning of the borrowed noun opens up for a transition into a verb via verbal roots: if you borrow the word *telephone*, it seems natural to want to express the action and verb *to telephone* for instance. I believe semantic reasons cause the morphological integration of 2b.

#### 4.5 Category 2c: loanwords with a verbal root and a broken plural

This category represents the words found in the corpus that have undergone the most complete form of integration into the morphology of MSA as they possess the two characteristics of category 2a and 2b together namely both a verbal root and a broken plural form. Their integration into the MSA morphology is thus complete both from a verbal and from a nominal point of view. The different derivatives that these words have will be explored below.

The first word is فيلسوف *faylasūf* ‘philosopher’ which has the root فلسف *falsafa* ‘to philosophize, reflect philosophically (s on something)’. It also exists in stem II *tafalsafa* ‘to philosophize; to pretend to be a philosopher’. Other words sharing the same root include فلسفة *falsafa* ‘philosophy’ (present in a lot of compounds), the nisba adjective فلسفي *falsafī* ‘philosophic(al)’ and two participles مفلسف *mufalsif* ‘philosopher’ and متفلسف *mutafalsif* ‘philosophaster, philosopher’. The broken plural is similar to the one of class iii: CvCCvC(at) -> CaCāCiCat with the difference that the first vowel in singular is long and not short, فيلسوف *faylasūf* ‘philosopher’ -> pl. فلاسفة *falāsifa*.

Last but not least we have the word شيطان *šayṭān* ‘Satan’ that comes from biblical Hebrew, according to the French and English etymological dictionaries. The word does not figure in Rolland (2014) which is not surprising as the word actually is of Semitic origin. It figures in the dictionary under the quadrilateral root شيطن *š-y-t-n* which in stem II *tašayṭana* means ‘to behave like a devil’. The broken plural form follows class iii CvCCvC(at) -> CaCāCiC: شيطان *šayṭān* ‘satan’ -> pl. شياطين *šayāṭīn*. شيطان *šayṭān* has several meanings: ‘Shaitan, Satan, devil, fiend’. There is a nisba adjective شيطاني *šayṭānī* ‘satanic, devilish, fiendish; demonic, demoniac, hellish, infernal’ as well as the word شيطنة *šayṭana* ‘devilry, villainy, dirty trick’.

The words in category 2c are then productive quadrilaterals. These words both fit an MSA pattern and have semantic reasons which can cause their advanced level of integration.

#### 4.6 Summary table: loanwords and the distribution of their MSA morphology

The sign “x” stands for existing features. Assumed forms are marked with “(x)”. Forms that do not exist or have not been encountered have the sign “-”. Adjectives have been excluded from the list.

Table 4

Loanword number	MSA, transcription	English translation	Suffixal fem. pl.	Broken pl.	Verbal root
1	بورنو <i>būrnū</i>	porn	-	-	-
2	تيلوب <i>tiyūlīb</i>	tulips	-	-	-
3	سيرك <i>sīrk</i>	circus	-	-	-
4	كنار <i>kanār</i>	canary	-	-	-
5	كولونيل <i>kūlūnīl</i>	colonel	-	-	-
6	نايلون <i>nāylūn</i>	nylon	-	-	-
7	بطاطا <i>baṭāṭā</i>	potato	(x)	-	-
8	بيرة <i>bīra</i>	beer	(x)	-	-
9	جغرافيا <i>juġrāfiyā</i>	geography	(x)	-	-
10	جينز <i>jīnz</i>	jeans	(x)	-	-
11	دش <i>duš</i>	shower	(x)	-	-
12	سينما <i>sīnamā</i>	cinema	(x)	-	-
13	كولونيا <i>kūlūniyā</i>	cologne	(x)	-	-
14	هستيريا <i>histiriya</i>	hysteria	(x)	-	-
15	ألبومات <i>’albūmāt</i>	photo albums	x	-	-
16	الالكترونيات <i>’ilēktrūnāt</i>	electronics	x	-	-
17	أوبرا <i>’ubrā</i>	opera	x	-	-
18	باص <i>bāṣ</i>	bus	x	-	-
19	بروجكتورات <i>brūjiktūrāt</i>	projector	x	-	-
20	بلكون <i>balkūn</i>	balcony	x	-	-
21	بلوزات <i>blūzāt</i>	blouses	x	-	-
22	بنطلون <i>banṭalūn</i>	trousers	x	-	-
23	ديكور <i>dīkūr</i>	set	x	-	-
24	روتين <i>rūtīn</i>	routine	x	-	-
25	رولات <i>rūlāt</i>	rolls	x	-	-
26	سكرتيرة <i>sakritīra</i>	secretary	x	-	-
27	غاز <i>gāz</i>	gas	x	-	-

Loanword number	MSA, transcription	English translation	Suffixal fem. pl.	Broken pl.	Verbal root
28	كاميرا <i>kāmīrā</i>	camera	x	-	-
29	كريمات <i>krīmāt</i>	creams	x	-	-
30	كلسون <i>kalsūn</i>	underpants	x	-	-
31	كمبيوتر <i>kumbiyūtir</i>	computer	x	-	-
32	موسيقى <i>mūsiqā</i>	music	x	-	-
33	كرتون <i>kartūn</i>	cardboard	x (less common)	x	-
34	كابلات <i>kābilāt</i>	cables	x	x	-
35	شال <i>šāl</i>	scarf	x	x	-
36	أفلام <i>ʿaflām</i>	films	-	x	-
37	بركان <i>burkān</i>	volcano	-	x	-
38	دكتور <i>duktūr</i>	doctor	-	x	-
39	دينار <i>dīnār</i>	dinar	-	x	-
40	روب <i>rūb</i>	robe	-	x	-
41	سجائر <i>sajāʿir</i>	cigarettes	-	x	-
42	فرن <i>furn</i>	oven	-	x	-
43	قنديل <i>qindīl</i>	lamp	-	x	-
44	كواليس <i>kawālīs</i>	wings	-	x	-
45	تلفزيون <i>tilifziyūn</i>	television	x	-	x
46	تليفون <i>tilīfūn</i>	telephone	x	-	x
47	أسفلت <i>ʿasfalt</i>	asphalt	-	-	x
48	إسمنت <i>ʿismant</i>	cement	-	-	x
49	صابون <i>šābūn</i>	soap	-	-	x
50	شيطان <i>šaytān</i>	Satan, devil	-	x	x
51	فيلسوف <i>faylasūf</i>	philosopher	-	x	x

## 5. Discussion

The view chosen in this thesis for observing loanwords may seem somewhat simplistic and unidimensional. First of all it is based on the loanword as a written word and in MSA. This excludes everything that concerns a more oral perspective on loanwords- their pronunciation; the use, integration and forms of loanwords in a colloquial context, everything that has to do with social attitudes towards loanwords etc. There is no doubt that borrowing occurs far more frequently in the colloquial dialects than in MSA, that the oral language is the meeting point for borrowing and the typical scene for language contact and that this might influence the morphological integration heavily. On a brighter side, this means that there is room for a lot of studies focusing on loanwords in spoken Arabic and that the material should not be difficult to find.

However, we should not ignore the fact that a written kind of borrowing exists almost on its own, with strong bonds to oral borrowing. It goes without saying that for Arabic both MSA and the dialects influence each other when it comes to loanwords: the stress and pronunciation of a loanword may well affect how loanwords are integrated in MSA. Spelling and morphological processes applied to the loanwords in MSA can of course color the oral use. A certain regional variation can also be expected due to the multifaceted nature of the Arabic speaking world, where the spoken language is not the same everywhere and presumably even MSA cannot be entirely uniform in its use and integration of loanwords.

The point here is that borrowing in MSA is more complex than what might be suggested by this thesis. The integration of loanwords in MSA morphology is part of a bigger picture with far more parameters than the ones considered here. The chain of reactions is still somewhat a mystery when it comes to crossing the border between words with solid stems and formal roots. Indeed there are words in category 1 that could very well fit in one of the broken plural classes but that for one reason or the other are still considered to be solid stems: why is the plural of *بلكون* *balkūn* 'balcony' *بلكونات* *balkūnāt* when it fits perfectly the pattern of class iii *CvCCv̄C(at)* -> *CaCāCīC*: *بلكون* *balkūn* 'balcony' -> pl. *\*بالاكين* *balākīn*? And there are far more examples of words from category 1 that without any change at all or with minor ones could be brought to a more advanced level of morphological integration. Yet another factor that should be considered is that the more *nominal* the word, the greater the likelihood that it will remain in category 1, as the other categories, and especially category 2b and 2c, display a *verbal* root. It is almost impossible to derive any denominal verbs from the words of category 1. The borrowed words from SAE are nominal but the roots created in MSA are verbal.

The root and pattern system of MSA is a key to the understanding of some of the processes involved in the integration of loanwords into MSA. Even though the solid stems of category 1 represent by far the largest part of the loanwords in the corpus, it is still remarkable that all the other loanwords from categories 2a, 2b and 2c can be mapped to a root and pattern of MSA and that it is possible to dissect both patterns and roots in these words as if they were native Arabic words. It could also be proposed that when speakers want to force a foreign word into their language, they can do this regardless of the typological distance between the languages involved, the words in groups 2a-2c being a good illustration of this.

Of course there are restrictions on a purely formal approach to loanword integration as many other factors might influence the degree of integration and the reason for separating loanwords into different layers of integration. These might include the time elapsed since the borrowing took place (the age of the loanword), it seems logical to believe that older loans have a higher chance of being integrated morphologically, whereas newer ones might be treated as solid stem words. Examples of that might be the level of integration of the word شيطان *šayṭān* 'Satan', older and more morphologically integrated versus كمبيوتر *kumbiyūtir* 'computer', newer and less integrated. The frequency of use might also play an important role, the more the loanwords are used, the more adapted. The level of bilingualism or monolingualism of the speaker community might influence as well. Patterns could eventually be found if a mapping of the source languages were conducted, it may be that some source languages are easier to integrate than others. Social and attitudinal factors such as the prestige associated with the loan and the social situation of the speaker could probably also contribute to incorporate some borrowed elements further morphologically than others. Finally, we have already mentioned that semantics plays an important role when it comes to deciding the number of a certain word and the relevance of doing so. The semantics of the borrowed noun most certainly decides which loanwords will be analyzed as having verbal roots, generating the production of a verb in MSA. A mapping of the semantic domains covered by the loanwords and a comparison with their level of integration would surely be of relevance as well.

The use of long vowels reduces the risk of homophones, as all those vowels are explicitly expressed, thus reducing possible confusion regarding the word referred to. The use of quadriliteral patterns or quadriliteral consonantal roots have a similar effect as there are few native words in Arabic that are quadrilaterals (and the extra root thus facilitates the identification of the word in question) whereas trilaterals are very common in MSA and may create homophony.

Loanword adaptation to the RL is not uncommon. Most authors believe that for a loanword to survive and to be used in the RL, there has to be at least some adaptation to the latter (Haspelmath 2009: 42, see also Matras 2009: 172-173). The present work has hopefully shown that loanwords form a broad spectrum of integration: we have loanwords that are barely integrated such as those of category 1 and loanwords that are integrated as witnessed by their use of an RL pattern as in category 2a. Then we have loanwords that are integrated by way of an RL grammatical structure (category 2b) and loanwords that use both a pattern and a grammatical structure of the RL, namely category 2c. This makes the border between loanwords and native words even fuzzier, and raises problems for a general definition of loanwords.

Complex situations arise when one confronts the findings of this study with the research of Haspelmath (2009). If the rule of analyzability were to be followed this would mean that the only words that would count as loanwords for Haspelmath would be those of category 1. The others, which can be analyzed in the RL, would not count. However, does it make sense to exclude the word *فلم* *film* from being a loanword because it has an RL pattern, even if this pattern is not visible from the structure of the word itself but only together with its plural form? Does analyzability always exclude a word from the loanword category? Maybe it is possible to say that broken plural patterns and root affiliation involve analyzability at a different level than the purely morphological one, due to their added abstract character.

We should also keep in mind that root attribution is somewhat a matter of grammatical interpretation, especially when it comes to loanwords. Wehr has been my canon regarding root affiliation of the loanwords of the corpus, but it is of course possible to have alternative interpretations. The identification of a loanword root can be done in many different ways. The masculine word *ديكور* *dīkūr* ‘set’ (or ‘décor’), part of category 1, could perfectly well belong to the root *دكر* *d-k-r*, but such an interpretation has not been made. For some of the words in the corpus, it is not unthinkable that they might evolve into a more advanced stage of morphological integration in the future. This would presumably make the general definition of loanword yet more problematic.



## 6. Conclusion

The results obtained from the identification of loanwords through the novels, the elimination of some words, and a subsequent formal and morphological analysis of the selected words with Wehr, Rolland, other dictionaries, and AC are clear.

We can identify four different levels of morphological integration, an important border separating category 1 (having solid stems) from categories 2a, 2b and 2c (having roots). There are even subtler differences between loanwords where formal roots can be construed by way of a broken plural pattern (2a), where a clear verbal root has been assigned in the dictionary (2b), and the combination of the two previous cases: an existence of both a verbal root and a broken plural (2c).

More concretely category 1 is characterized by feminine suffixal plurals in *-āt*, a spelling where the vowel sounds are represented by long vowels and a general inability to adapt to the MSA root and pattern system. Category 2a consists of words where the construal of formal roots made possible by the application of broken plural patterns gives rise to more trilateral and fewer quadrilateral roots, often by using a restricted number of inflection patterns. Category 2b, loanwords with a verbal root, has mostly quadrilateral members and category 2c with verbal roots and broken plurals consists exclusively of quadrilaterals.

So to answer the question of how morphological integration of loanwords occurs in MSA we can say that the loanword is integrated into one of these four levels, along with their respective characteristics. This morphological integration is surely linked to other factors as well such as the age of the loanword and the frequency of usage.

Why loanwords are integrated differently, that is, why some are barely integrated except for the common integration undergone by all loanwords, whereas others are fully integrated verbally and nominally, is more difficult to say. Pattern congruity is the main reason provided in this thesis, i.e. the structure of the borrowed word fits a pre-existing inflection class or pattern of MSA such as the one for broken plurals. Loanwords that formally fit such a pattern will be integrated nominally and there is a possibility of extracting roots. Once a loanword can be analyzed in terms of roots, and even more if it is productive, the integration can be said to be complete. The only explanation to why loanwords are integrated morphologically with varying success provided here is based on the form of the loanword and the type of patterns available in MSA that it could be mapped to. Semantic aspects of the borrowed noun might also influence the probability of acquiring verbal roots (category 2b, 2c).

The existence of a verbal root in a loanword would then be both morphological evidence of a high degree of integration and the result of semantics. Formal roots as in category 2a have no corresponding verbal meaning in Arabic and are unproductive, wherefore the loanword retains a certain foreign character. On the contrary, truly productive roots, such as those in category 2b and 2c, enable the loanword to be fully incorporated into the Arabic language. So maybe semantics overrules morphophonology in a certain way.

The spelling of unintegrated loanwords of category 1 is an area that would require more research and a more thorough investigation than the speculations made in the present thesis. The original hypothesis that loanwords written with several long vowels are more difficult to integrate than loanwords with short vowels appears to be correct, but the reasons behind this finding, except for a mismatch between spellings with long vowels to show vowel quality and the presence of very few patterns containing several long vowels, remain unclear. A study focusing entirely on the spelling and syllable structure of loanwords in correlation with their level of integration would yield very interesting results.

On the whole, at a very general level, the loanwords in MSA encountered in the corpus fit with the definitions provided in the theory section about borrowing and loanwords (see 2.2). However, the findings of this thesis shed light on some problematic aspects of the general terminology and classification systems of loanwords of Haspelmath (2009) regarding the analyzability of loanwords in the RL. It can still be debated where on the morphological scale of integration loanwords in MSA cease to be loanwords if analyzability is included as a factor. This rule had to be excluded for the thesis. The border could be drawn once loanwords can fit into a broken plural pattern, or when they are linked to a verbal root or when they are fully integrated with both a broken plural and a verbal root.

It can also be argued that loanwords possessing a broken plural pattern are not analyzable in their singular form, as the pattern is not morphologically visible there. Those loanwords are analyzable on a mental level only if we see the singular form, while knowing the broken plural one and if we, at the same time, can picture the pattern used for going from the singular to the plural. If the loanword appears in its broken plural form it might carry more information, but we would still need some mental processing to get to the singular form and to represent the pattern. The words that have a broken plural can be analyzed to a certain extent, but is it the same type of analysis as the more instantaneous recognition of the morphs forming English compounds like *ethnography*, *ethnology* etc. (see Haspelmath 2009: 37)?

The somewhat arbitrary character of root attribution of loanwords is also problematic for deciding whether a loanword is analyzable from an MSA standpoint. Loanwords can have different roots depending on the interpretation made, as for the root of صابون *ṣābūn* ‘soap’ that could be either trilateral or quadrilateral or nouns with a CVC structure for instance. Some loanwords that lack roots could surely evolve into having roots in the future. It is always possible to identify roots in a word, but it is another story to anchor roots in speakers and in the language itself. On the other hand, it would seem that a verbal root can be construed by speakers of the RL when the need is strong enough, speakers’ agentivity thus overruling the pattern congruity constraint.

The field of Arabic etymology would benefit from the publication of a complete etymological dictionary in MSA. Luckily such a project is underway, lead by Stephan Guth, Professor of Arabic at the University of Oslo (see <http://www.hf.uio.no/ikos/english/research/news-and-events/news/2014/a-treasure-trove-of-arabic-terms.html>). The research will hopefully be available soon as a comprehensive electronic database, *EtymArab*, where it will be possible to trace words back in time to study their evolution and also to access all the knowledge that could be expected from an ordinary etymological dictionary.

An advantage with studying morphological integration between languages that differ a lot morphologically and typologically is that it is easy to distinguish the contribution of the separate languages. By tradition, many studies have been conducted on closely related languages, often from the SAE language family, which might lead to less tangible results as it might be difficult to separate influence from the SL from internal development within the RL.

There is no doubt that the RL has the upper hand on loanword integration. As I mentioned in the beginning, Winford states that loanwords “have to be adapted to the syntax and morphology of the RL, particularly if it has rules involving categories like case, number, gender, agreement, and the like” (Winford 2010: 173). Some of these parameters have been studied in this work and the focus has been on the MSA morphology that these loanwords possess. “In all cases, borrowed items are manipulated so that they conform to the structural and semantic rules of the RL. This is the hallmark of borrowing under RL agentivity” (Winford 2010: 175). This has also been very obvious in the present thesis.

Loanwords are an interesting area of research as a deeper understanding of loanwords can shed light on specific behaviors of the different languages involved, things that we might not pay any attention

to if it were not for the contact with another language that works in a different way. Loanwords are a good way of studying distinctive features of languages. Loanwords can thus highlight aspects of the different languages between which the borrowing takes place but can also contribute to the general knowledge about language contact, a fascinating area of linguistic study encompassing amongst others multilingualism, language convergence, code-switching, and language mixing.

## 7. Bibliography

- Badawi, El-Said M., Carter, Michael G. & Gully, Adrian (2004). *Modern written Arabic: a comprehensive grammar*. London: Routledge
- Haspelmath, M. (2009) "II. Lexical borrowing: concept and issues". In Haspelmath, Martin & Tadmor, Uri (red.), *Loanwords in the world's languages: a comparative handbook*, 35-54. Berlin: Mouton De Gruyter
- Haywood, John Alfred & Nahmad, Hayim Musa (1993). *A new Arabic grammar of the written language*. 2. ed London: Lund Humphries
- Heath, Jeffrey (1989). *From code-switching to borrowing: foreign and diglossic mixing in Moroccan Arabic*. London: Kegan Paul International
- Holes, Clive (2004). *Modern Arabic: structures, functions, and varieties*. Rev. ed. Washington, D.C.: Georgetown University Press
- Matras, Yaron (2009). *Language contact*. Cambridge: Cambridge University Press
- Matras, Yaron & Sakel, Jeanette (red.) (2007). *Grammatical borrowing in cross-linguistic perspective*. New York: Mouton de Gruyter
- Newman, Daniel L. & Husni, Ronak (red.) (2008). *Modern Arabic short stories: a bilingual reader*. London: Saqi
- Owens, Jonathan (2013). "A house of sound structure, of marvelous form and proportion: an introduction". In Owens, Jonathan (red.). *The Oxford handbook of Arabic linguistics*, 1-23. Oxford: Oxford University Press
- Pereltsvaig, Asya (2012). *Languages of the world: an introduction*. Cambridge: Cambridge University Press
- Ratcliffe, Robert R. (2008). "Number". In Versteegh, C. H. M. & Eid, Mushira (red.). *Encyclopedia of Arabic language and linguistics*. Vol. 3: 439-447, Lat-Pu, Leiden: Brill
- Reichmuth, Philipp (2009). "Transcription". In Versteegh, C. H. M. & Eid, Mushira (red.). *Encyclopedia of Arabic language and linguistics*. Vol. 4: 515-520, Q-Z. Leiden: Brill
- Rolland, Jean C. (2014). *Les mots de l'arabe moderne d'origine non sémitique*. Diss.
- Ryding, Karin C. (2005). *A reference grammar of Modern Standard Arabic*. Cambridge, UK: Cambridge University Press
- Sa'id, Majed F. (1967). *Lexical innovation through borrowing in Modern Standard Arabic*. Diss.
- Sakel, Jeanette (2007). "Types of loan: Matter and pattern". In Matras, Yaron & Sakel, Jeanette (red.) *Grammatical borrowing in cross-linguistic perspective*, 15-31. New York: Mouton de Gruyter
- Smeaton, B.H. (1973). *Lexical expansion due to technical change as illustrated by the Arabic of Al Hasa, Saudi Arabia*. Bloomington: Indiana university publications. Diss.
- Tadmor, U. (2009). "III. Loanwords in the World's Languages: Findings and results". In Haspelmath, Martin & Tadmor, Uri (red.), *Loanwords in the world's languages: a comparative handbook*, 55-75. Berlin: Mouton De Gruyter.

Versteegh, C. H. M. & Eid, Mushira (red.) (2009). *Encyclopedia of Arabic language and linguistics. Vol. 1-4*. Leiden: Brill

Wehr, Hans (1952-1959). *Arabisches Wörterbuch für die Schriftsprache der Gegenwart*. Leipzig.

Wehr, Hans & Cowan, J. Milton (1994). *A dictionary of modern written Arabic (Arabic-English). 4th ed. considerably enl. and amended by the author*. New York: Spoken Language Services

Winford, Donald (2010). "Contact and borrowing". In Hickey, Raymond (red.). *The handbook of language contact*, 170-187. Chichester: Wiley-Blackwell

Electronic references:

<http://arabiccorpus.byu.edu>

<http://www.cnrtl.fr>

<http://www.hf.uio.no/ikos/english/research/news-and-events/news/2014/a-treasure-trove-of-arabic-terms.html>

<http://www.oed.com>

## Appendix

1. Loanwords without any root affiliation, lacking verbal root and broken plural (in alphabetical order):

MSA, transcription	English translation	Suffixal fem. pl.	Wehr	Rolland	Remarks
ألبومات <i>albūmāt</i>	photo albums	yes	p. 28	-	-
الالكترونيات <i>'iliktrūnāt</i>	electronics	yes	p. 29	p. 107	Also الكترونيات <i>'iliktrūnīyāt</i> .
أوبرا <i>'ūbrā</i>	Opera	yes	p. 42	-	-
باص <i>bās</i>	bus	yes	p. 50	-	-
بتروولية <i>bitrūliyya</i>	petrochemical	yes	p. 52	p.120	Nisba adj., plural form irrelevant.
بروجكتورات <i>brūjiktūrāt</i>	projector	yes	-	-	In AC. In pl. in the novel.
برونزية <i>brūnziyya</i>	tanned	yes	p. 69	p. 127	Metaphor; pl. irrelevant.
بطاطا <i>baṭāṭā</i>	potato	probably	p. 77	-	Some instances of fem. Suffixal pl. in AC. Ends in <i>-ā</i> .
بلاستيكي <i>blāstīkīy</i>	plastic	yes	p. 87	p. 134	Nisba adj., pl. irrelevant.
بلكون <i>balkūn</i>	balcony	yes	p. 90	p.136	-
بلوزات <i>blūzāt</i>	blouses	yes	p. 92	p. 137	-
بنطلون <i>banṭalūn</i>	trousers	yes	p. 94	p. 140	-
بورنو <i>būrnū</i>	porn	-	-	-	Word found in AC.
بيرة <i>bīra</i>	beer	probably	p. 103	p. 146	Also بيرا <i>bīrā</i> .
تيليب <i>tiyūlīb</i>	tulips	-	p. 121	p. 157	-
جغرافيا <i>juḡrāfiyā</i>	geography	probably	p. 178	p. 167	Also جيوغرافيا <i>jiyūḡrāfiya</i> . Few occurrences of suffixal fem. pl. in AC. Ends in <i>-ā</i> .

MSA, transcription	English translation	Suffixal fem. pl.	Wehr	Rolland	Remarks
جينز <i>jīnz</i>	jeans	probably	-	-	Word common in AC. <i>Baṅṭālūn, -āt</i>
دش <i>duš</i>	shower	probably	p. 325, 348	p. 179	Variants: دوش <i>dūš, -āt</i> (Wehr) Rolland: دشّ <i>dušš</i> .
ديكور <i>dīkūr</i>	set	yes	p. 352	p. 184	-
روتين <i>rūtīn</i>	routine	yes	p. 422	-	Suffixal fem. pl. in AC.
رولات <i>rūlāt</i>	rolls	yes	-	-	Somes instances in AC.
سكرتيرة <i>sakritīra</i>	secretary	yes	p. 487	p. 205	Animate fem.
سيرك <i>sīrk</i>	circus	-	p. 522	p. 202	سرك <i>sirk</i> in Rolland.
سينما <i>sīnamā</i>	cinema	probably	p. 524	p. 212	Ends in <i>-ā</i> .
غاز <i>gāz</i>	gas	yes	p. 777	p. 233	-
كاميرا <i>kāmīrā</i>	camera	yes	p. 946	p.257	-
كريمات <i>krīmāt</i>	creams	yes	p. 965	p. 269	Spelled كريم <i>krīm</i> and كريمة <i>krīma</i> (in sg.)
كلسون <i>kalsūn</i>	underpants	yes	p. 980	p. 256	Normally for men but used for a girl in the text.
كمبيوتر <i>kumbiyūtīr</i>	computer	yes	-	-	Suffixal fem. pl. in AC.
كنار <i>kanār</i>	canary	-	p. 986	p. 273	كناري <i>kanārīy</i> in Rolland and Wehr.
كولونيا <i>kūlūniyā</i>	cologne	probably	p. 993	p. 277	Ends in <i>-ā</i> .
كولونيل <i>kūlūnīl</i>	colonel	-	p. 993	-	Animate, semantically m.



MSA, transcription	English translation	Suffixal fem. pl.	Wehr	Rolland	Remarks
منغولي <i>munḡūlīy</i>	with Down's Syndrome	yes	p. 1088	-	'Mongolian' (in Wehr). Nisba adj., pl. irrelevant.
موسيقى <i>mūstīqā</i>	music	yes	p. 1093	p. 297	Suffixal fem. pl. in AC. Ends in -ā.
موسيقية <i>mūstīqiyya</i>	music	yes	p. 1093	p. 297	Nisba adj., pl. irrelevant.
موف <i>mūv</i>	purple	-	-	-	New letter for /v/--> ف Adjective, pl. probably irrelevant.
نايلون <i>nāylūn</i>	nylon	-	p. 1190	p. 305	نيلون <i>nīlūn</i> in Wehr and Rolland.
هستيريا <i>histiriya</i>	hysteria	probably	p. 1205	p. 307	هستيريا <i>histiriya</i> (in Rolland and Wehr) and هيسستيريا <i>hīstiriya</i> in Wehr. Ends in -ā.

Total category 1: 38 words

2a. Loanwords with a broken plural form that provides a formal root (in alphabetical order):

Singular	Broken plural	English translation	Class	Root	Wehr	Rolland
بركان <i>burkān</i>	براكين <i>barākīn</i>	volcano	3	بركن b-r-k-n	p. 67	p.124
دكتور <i>duktūr</i>	دكاترة <i>dakātira</i>	doctor	3	دكتر d-k-t-r	p. 333	p. 180
دينار <i>dīnār</i>	دنانير <i>danānīr</i>	dinar	3/4	دندر d-n-r دننر d-n-n-r	p. 354	p. 185
روب <i>rūb</i>	أرواب <i>'arwāb</i>	robe	1	روب r-w-b	p. 422	p. 190
سجارة <i>sijāra</i>	سجائر <i>sajā'ir</i>	cigarette	4	سجر s-j-r	p. 462-463	p. 211
شال <i>šāl</i>	شيلان <i>šīlān</i> ; also suffixal fem. pl.	scarf	1	شيل š-y-l	p. 525, 582	p. 213-214
فرن <i>furn</i>	أفران <i>'afrān</i>	oven	1	فرن f-r-n	p. 832	p. 240
فلم <i>film</i> , فيلم <i>ftlm</i>	أفلام <i>'aflām</i>	film	1	فلم f-l-m	p. 852, 862	p. 243
قنديل <i>qindīl</i>	قناديل <i>qanādīl</i>	lamp	3	قندل q-n-d-l	p. 927	p. 258
كابل <i>kābil</i>	كوابل <i>kawābil</i> ; and suffixal fem. pl.	cable	5	كبل k-b-l	p. 944	p. 262
كرتون <i>kartūn</i> , كارتون <i>kārtūn</i>	كراتين <i>karātīn</i> ; less common suffixal fem. pl.	cardboard	3	كرتن k-r-t-n	p.945, 959	p. 172
كوليس <i>kūlīs</i>	كواليس <i>kawālīs</i>	wings (‘coullisse’)	4	كلس k-l-s	p. 993	p. 277

Total category 2a: 12 words

2b. Loanwords with a verbal root (in root order):

MSA, transcription	English translation	Root	Plural	Wehr	Rolland
تلفزيون <i>tilifiziyyūn</i> تليفزيون <i>tilifiziyyūn</i>	television	تلفز <i>talfaza</i> 'to televise, transmit by television'	Fem. suffixal pl.	p. 116	p. 153
تليفون <i>tilifūn</i> , تلفون <i>tilifūn</i>	telephone	تلفن <i>talšana</i> 'to telephone'	Fem. suffixal pl.	p. 116	p. 153
أسفلت <i>'asfalt</i>	asphalt	سفلت <i>saflata</i> 'to cover with asphalt, to asphalt'	-	p. 20, 482	p. 101
إسمنت <i>'ismant</i>	cement	سمنت <i>samnata</i> 'to cement'	-	p. 21, 503	p. 102-103
صابون <i>šābūn</i>	soap	صبن <i>š-b-n II</i> 'to soap, rub with soap' (Wehr); صوبن <i>šawbana</i> 'savonner' (Rolland)	-	p. 586	p. 220

Total category 2b: 5 words

2c. Loanwords with verbal root and broken plural (in root order):

Singular	Broken plural	English translation	Class	Root	Wehr	Rolland
شيطان <i>šayṭān</i>	شياطين <i>šayāṭīn</i>	Satan, devil	3	شيطان <i>š-y-t-n II</i> 'to behave like a devil'	p. 581	-
فيلسوف <i>faɣlasūf</i>	فلاسفة <i>falāsifa</i>	philosopher	3	فلسف <i>falsafa</i> 'to philosophize, reflect philosophically'	p. 850, 862	p. 242

Total category 2c: 2 words