The Arabic Origins of Numeral Words in English and European Languages

Zaidan Ali Jassem

Department of English Language and Translation, Qassim University,

P. O. Box 6611, Buraidah, KSA

Tel: 966-6-385-2338   E-mail: zajassems@gmail.com

Abstract

The aim of this paper is to the examine the genetic relationship between all numeral words in Arabic and English primarily as well as other European languages such as German, French, Greek, Latin and even Sanskrit secondarily. Contrary to long-held beliefs and views in Western comparative historical linguistics in which Arabic and English, for example, are classified as members of different language families, it shows how these numerals are related to and derived from one another, with Arabic being the end origin perhaps. The numeral words have the same or similar forms and meanings with slight phonetic changes as a result of normal linguistic evolution at the phonetic, morphological and lexical levels.

Keywords: Numerals, Arabic, English, German, Comparative linguistics, Lexical root theory

1. Introduction

Historical comparative linguists use numerals, especially low ones, to establish genetic relations between languages as they are part of the universal core or basic vocabulary which resists borrowing across languages (Campbell 2004: p. 126; Crowley 1997: p, 171; Pyles and Algeo 1993: pp. 76-77). The core numerals are usually the low numerals (Pyles and Algeo 1993: pp. 76-77) which vary in number according to the word list used. In Swadesh's 100-word list, only the numerals *one* and *two* occurred (Campbell 2006: pp. 201-202) whereas in his 200-word list (Crowley 1997: p. 174) the numerals 1, 2, 3, 4, 5, 7, 8, 10, 20 and 100 were included; thus 6 and 9 were not on the list. These numerals have mostly similar or identical forms in all European or Indo-European languages of all branches like the Germanic family (e.g., English, German), the Italic (e.g., French, Italian), the Hellenic (e.g., Greek), the Slavic (e.g., Russian) and the Indic (e.g., Sanskrit) (for a survey, see Crystal 2010: p. 308; Campbell 2006: pp. 190-191).

According to all comparative historical linguists, Arabic and English belong to entirely different language families: one Semitic and one Germanic (e.g., Crystal 2010; Viney 2008; Kirkpatrick 2007; Campbell 2006; Pyles and Algeo 1993; Ruhlen 1987, 1994). Therefore, Arabic numerals are not genetically related to those of English or any Indo-European family. This paper will prove that they really are, with Arabic being the end origin of them all from one to trillion. The paper has five sections: section one is introductory, section two introduces the numeral data, section three deals with data analysis, section four describes the results, and section five is discussion and conclusion.

2. The Data: Numeral Words in Arabic and English

The data consists of all the numerals, both low and high, in Arabic and English in the main as well as other European languages, which will be examined to confirm not only their common genetic origin but also their Arabic origin as well. In fact, the main reason for including numerals from other European languages here is to trace linguistic development so as to be able to resolve linguistic matters and set up relationships accurately. All these Arabic and English numerals are included in the dialogue below.

Saphia: Can you count from one to ten in English, Sally?

Sally: Yes, I can count from one even to trillion. One, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen… nineteen, twenty, twenty one, thirty … ninety, a hundred, a thousand, a million, a billion, and a trillion. Or first, second, third, fourth, fifth, sixth, etc.

Saphia: Very good. Can you count that in Arabic?

Sally: Yes, I can. waa2id, ithnan, thalathat, arba3at, khamsat, sittat, sab3at, thamaniat, tis3at, 3ashrat, a2ad 3ashrat, ithnata 3ashar, thalata 3ahar …. 3ishreen, waa2id wa 3ishreen, thalatheen …. tis3een, miyat, alf, alf alf (malyon), balyon, trilyon. Or 2aadi (or awwal), thani, thalith, raabi3, khamis, sadis, etc.

Saphia: Excellent. Can you count in French and German also?

Sally: Yes. I can also do that in Latin, Greek, Sanskrit and Russian because all such numbers are alike in general.

Mandy: You're stunningly right and bright.

As can be seen from the passage, Arabic and English numerals are two types: cardinal and ordinal numbers. However, there are a few morphological and lexical differences between them. First, unlike English numerals, the Arabic numerals 1-10 have masculine and feminine forms, the latter of which have the suffix *–at*, which is the normal way of saying them in isolation. However, the normal way for saying the first two numerals is the masculine form, whose feminine counterparts are *wa2idat* or *i2da* and *thintan* or *ithnatan* respectively. Again the same feminine suffix is used, albeit as an infix in *ithnatan*. Secondly, both the Arabic and English numerals 11-19 are in general a combination of the relevant cardinal and *ten* except for the first two in English. Thirdly, the Arabic numerals 20-90 have the plural suffix *–een* (or *–oon* depending on case) attached to the masculine form while English numerals attach *–ty*. Finally, ordinal numerals in English affix *–th* except for *first, second* (and *third*); Arabic ordinals apply internal and external vocalic changes except for *awwal* (first), which is a different or new word. At this juncture, it has to be noted that all initial Arabic (and English) vowels regularly start with a glottal stop when said in isolation, which is usually dropped or turned into /w/ in actual, everyday speech in Arabic. In medial and final postvocalic positions, it is often elided or pronounced /y/ as in *miyat* (hundred). For these reasons, the glottal stop is not shown in the transcription of *a2ad* (one), *awwal* (first) , *ithnan* (two), *arba3at* (four) and *miyat* (hundred) here.

3. Data Analysis

3.1 Theoretical Framework: Lexical Root Theory

The theoretical framework for analyzing the data will be the lexical root theory, which has been proposed here to tackle the situation at hand. Its main principles can be summarized as follows. The lexical root theory is so called because it is based on the lexical root of the word in examining genetic relationships between words such as *write* and *katab* (write). It has a principle or construct and four components. The principle is theoretical in nature which states that Arabic and European languages of all branches are not only genetically related but also are derived from Arabic in the end. The four components, which are all practical in essence, are (i) a procedural component, (ii) a semantic component, (iii) a form-meaning component, and (iv) a linguistic analysis component. These components are the steps to be followed in analyzing genetic relationships between lexical roots, which are described briefly below.

First, the procedural component deals with how to proceed with the analysis of words by (i) deleting affixes, (ii) using primarily consonantal roots, (iii) using semantic fields (numerals in the present case), and (iv) search for meaning. That is, in order to genetically relate *underwriting* or *uniqueness* to their Arabic cognates, for instance, they must be reduced to *write* and *one* first. Secondly, the semantic component looks at the semantic relationships between words like stability, multiplicity, divergence, convergence, shift, and change. Stability means the meanings of words have not changed; multiplicity denotes that words

might have two or more meanings; divergence signals that words have become opposites or antonyms of one another; convergence denotes that two similar words might have led to another one with which it shares a similar form and meaning; shift indicates that words have exchanged, switched or transferred their reference or sense within the same field or category; change means a new meaning developed. Thirdly, the form-meaning component examines the relationship between form and meaning as to whether words are similar in form and meaning, similar in form but different in meaning, or different in form but similar in meaning. Finally, the linguistic component considers the linguistic analysis of words in the way their phonetic, morphological, grammatical and lexical structure is responsible for any differences between the numerals. In the following analysis, all the above components will be utilized, though with different degrees of focus.

3.2 Method of Analysis

The method of studying and establishing the genetic relationship between the numerals in English and Arabic is comparative where every numeral word in English in particular and German, French, Greek, Latin, etc. in general will be compared with its counterpart in Arabic phonetically, morphologically and semantically. In addition, it is historical where language development is taken into consideration as numeral words might have changed, swapped or reversed their forms and meanings from one language into another altogether. The sources of such meanings are Arabic dictionaries (e.g., Ibn Manzour 1974; Ibn Seedah 1996) and English etymological dictionaries (e.g., Harper 2012) in addition to the author's knowledge of both Arabic as a mother tongue and English as a specialty. Indeed, it is almost impossible to relate words without knowing their origin, history, and meaning.

4. Results: The Arabic Origins of English Numerals

Relating English and Arabic numerals is not too difficult to make. Some are very obvious, straight and easy to relate; some are less so especially when the different forms of cardinal and ordinal numbers are taken into account; and some are very vague, indeed.

4.1 The Numeral One

In English, *one* is the cardinal number and *first* is the ordinal; in Arabic, *wa2id* or *a2ad* (one) is the cardinal and *2adi* (first) is the ordinal. As can be seen, these cardinals and ordinals do not seem to be related at all. However, they really are because Arabic has another commoner synonym for *2adi* which is *awwal* (one, first). That is, *one* is directly derived from the latter, commoner Arabic ordinal word *awwal* via two sound changes: (i) initial syllable deletion (i.e., /aw/-loss) or /?/-merger with /w/ as commonly happens in Arabic in similar cases such as *?akhkhir → wakhkhir* (go back) and *?irth → wirth* (inheritance) and (ii) the replacement of /l/ by /n/. In other words, the course of this phonetic change in English might look like this: *awwal → wal → wan (one)*.

The English words *unique(ness), unit, unity, united, Unitarian, unify, unification, unison, uni-, once, inch,* the indefinite articles *a* and *an* as well as *only, alone, lonely* are all directly related to the numeral *one* (Harper 2012) and, consequently, to *awwal* via similarity in meaning and form. All these words are either further simplifications or have fewer additions to the basic

root of *one* and/or *awwal*.

In Latin, French and German, similar forms to *one* are used such as *un(us)* in the first; however, different forms are used in Sanskrit and Russian, which are *eka* and *adin*. Thus, while *unus* derives from *awwal* above, *eka* and *adin* derive from the Arabic cardinal *wa2id* or *a2ad* (one). More precisely, in *eka* /2/, a voiceless fricative pharyngeal, became /k/ and /d/ was deleted as happens in certain coastal varieties of Syrian Arabic (e.g., Jassem 1993, 1994a-b) whereas in *adin* /2/ was deleted.

What about the ordinal number *first*? It is derived from the Arabic word *fard/fareed* (one/unique, single, person) in which /d/ either split into /s/ and /t/ or changed to /t/ with /s/ being an insertion. Alternatively, on the basis of its original meaning *earliest, preceding all others in time* (Harper 2012), it may be derived from *baakir, bikr* (early, first, virgin) in which /b/ turned into /f/ while /k/ split into /s & t/.

Finally, the more commonly used Arabic numeral *wa2id* (or *a2ad*) (one) led to the emergence of English *odd* (as in *odd* and *even* numbers) in which /?/, a voiceless glottal stop, and /2/, a voiceless fricative pharyngeal, were either deleted or merged into /w/. This derivation is further corroborated by the other meanings and derivatives of *odd* such as *odds, oddity,* and *oddness* (enmity, animosity, unfriendliness, strangeness, being one of a kind), all of which are cognates of the similar Arabic root *3adoo* (enemy) and related derivatives such as *3adawat* (n) (animosity, oddity) in which /3/, a voiced fricative pharyngeal, was deleted. Now one should be able to see how close *odd* to *3adoo* (enemy) and *a2ad* or *wa2id* (one) in sound and sense! They are similar to one another in form and meaning both. Thus they are real cognates.

4.2 The Numeral Two

In English, the words *two* (*dual, duality*, *duet, duo, double, duplicate*) and *twin* are related to one another and can be easily traced back to their Arabic source cognates. *Twin* is a phonetic simplification of Arabic *ithn(at)ain* (two) and/or *thintan* (two-fem.) in which /th/ became /t/ and /i/ backed or shifted to /w/. *Two* is a further phonetic corruption or reduction of *twin*, having lost its final /n/, which reappears in <u>twenty</u> (two/twin tens) (see 4.12 below). Or it could have evolved directly from the simpler ordinal Arabic numeral *thani* (second), in which /th & n/ merged into /t/. This is more plausible because in *ithn<u>an</u>* or *ithn<u>atan</u>*, *-an* is a dual suffix while *–at* is a feminine one. In short, the course of such sound change in English might be charted as *thintan → tintan → twintan → twin → twi* (two).

All European languages have similar forms such as *duo* in Latin, *dyo* in Greek, *zwēne* in Old High German, and *zwei* in Modern German. Thus their Arabic origins are obvious. The presence of /n/ in *zwēne* shows that *zwei* and *two* come from *ithn(at)an* in Arabic.

To these one can add *even* as in *odd and even numbers* and *twine/ intertwine*, both of which derive from the same Arabic source for *two*. That is, *even* comes from *ithnan, thani* in which /th/ became /v/ and lexical shift occurred as *even* indicates numbers which are multiples or divisibles of *two*. *Twine* derives from *thana* (bend, twine), a related derivative, in which /th/ turned into /t/. Notice how close the sequence *odd, even* to *wa2id, ithnan* are. Thus, if you were to count by using *odd, even* slowly and repetitively instead of the normal *one, two,* it's

highly likely that you would be understood as saying *wa2id (a2ad), ithnan* in corrupt Arabic, though.

As for *second* (in Latin, *secundus*), tracing it back to its Arabic cognate depends on its original meaning(s) in Latin: *following, favourable, assist, give support, 60th part of a minute* (Harper 2012). There are two formally similar but semantically different Arabic words, which combine all these meanings. The first is the feminine form of *two* in Arabic, which is *thintan* or *ithnatan* (two- fem.), and its ordinal *thaniyat* (second- fem., a time second, following, other), where /th/ split into /s/ and /k/ while /t/ turned into /d/. The second is *athna* (praise, thank, support), a related derivative. In fact, the initial, real, major clue for such a derivation is the presence of /-nd/ in *second* which corresponds to /-nt/ in the Arabic cognates. Thus, all the various forms are due to different phonetic and morphological changes. Such a sound change can be schematized as follows: *thintan → sintan → sindan → skindan → skind* (second).

4.3 The Numeral Three

The words *three, third* (also *tertiary, trio, tri-*) are very explicitly genetically related to their Arabic cognate *thalath(at)* (three) and related derivatives like *thulth, thalith* (1/3, third). In *three*, /l/ was replaced with /r/ whereas final /th/ was either lost or assimilated into the initial one as a result of simplification; however, /th/ reappears as /d/ in *third* (for more discussion see 4.12 below). To sum it up, the course of this sound change in English might look like *thalath → thala → thara → thra* (three).

As to other European languages, similar forms are used such as *tres* in Latin, *treis* in Greek, *trei* in German and *trois* in French. Similar phonetic processes happened, turning initial and final /th/ into /t & s/ each while /l/ into /r/.

4.4 The Numeral Four

Four is almost identical in both form and meaning to its Arabic cognate *arba3(at)* (four) and related derivatives *rub3, rabi3* (1/4, fourth) which underwent slight phonetic changes. First, the order of the sounds had been completely reversed; secondly, /b/ changed to /f/; finally, final /3/, a voiced fricative pharyngeal, was lost. In addition, the same Arabic cognate gave rise to *pair* in English which underwent reversal, /3/-deletion and lexical shift from four to two. The course of such sound change in English might look like *arba3 → bar3 → bar* (pair) → *far* (four).

From the Latin *quattuor* (in French *quatre*) and *tessares* or *tettares* in Greek came the English words *quart, quarter, quartet*, *quadrant, quadri-, quarantine* and *carat*. As they originally mean a *fourth* of something or something *cut* (Harper 2012), then their immediate Arabic cognate is *qeeraT, qarareeT* (pl.) (a quarter, a measurement unit) and its verbal root *qaraT* (cut), *qarTat* (n) via reordering in Latin and turning /q/, a voiceless uvular stop, into /s/ or /t/ in Greek; also /T/, the emphatic counterpart of /t/, changed to plain /t/. It is worth noting here that /q/ may be pronounced with all these variants in Arabic besides others, of course (e.g., Jassem 1993, 1994a-b).

4.5 The Numeral Five

The numeral *five* is directly derived from its Arabic cognate *khams(at)* (five), to which it does not seem to have any explicit genetic relationship. Phonetically and semantically speaking, it does, however. First, /kh/, a voiceless velar fricative, and /s/ merged into /f/, then /m/ assimilated into /v/. The course for such sound change in English might look like *khams → khamf → famf → fanf* (Modern German *fünf*) → *faf* (five). So the story is all about assimilation.

In Latin, French, and Geeek, *quinque, cinque* and *pente* are used in that order. All derive from the same Arabic cognate as well via different phonetic processes. While /m/ turned into /n/ in all and final /s/ became /k/ in the first two, /kh/ became /k/ in Latin but /s/ in French. In Greek, /kh/ underwent a further change, turning it into /p/ via /h/ or /w/ perhaps while /s & t/ merged into /t/. To sum it up, the course for such sound change in Latin and French might look like *khams → kams → kamk → kank* (*quinque* in Latin) → *sank* (*cinque* in French). For Greek, it might have proceeded like *khamsat → kham(th/f)at → famfat → fampat → pampat → panat* (*pente*).

4.6 The Numeral Six

The numeral *six*, which has slightly different spellings and pronunciations in German (*sechs*), French (*sees*) and Italian (*sestet*), has a very obvious Arabic origin in form and meaning: viz., *sit(tat)* or a related derivative like *suds, sadis* (1/6, sixth) in which /d/ or /t/ passed into /k/. Notice how close Arabic *sit* or *suds* is to French *sees* and Italian *sestet*. Schematically, such a sound change might have taken the following course in English: *sud*s → *suks* → *siks* (six).

As to the Latin form *six* and Greek form *hex*, both have the same story, in the latter of which /s/ turned into /h/.

4.7 The Numeral Seven

Like *six* above, the numeral *seven* is even still much closer to Arabic *sab3(at)* and related derivatives such as *sab3een* (seventy) in which /3/ merged with /b/ as /v/. In German *sieben*, /b/ is retained while /3/ is deleted (or turned into /n/ as happens in certain cases in Arabic dialects like *a3Ta* v. *anTa* (give)). In the author's dialect, this is normally the case (e.g., Jassem 1993, 1994a-b). So only the consonant /3/ had disappeared. In summary, the course of this sound change in English might be diagrammed as follows: *sab3at/sab3een → sabnat/sabeen → saveen* (seven).

In Latin, the form *septem* is used, in Sanskrit *sapta*, and in Greek *hepta*; all share the same story, in all of which /3/ was deleted while /s/ turned into /h/ in the last. Notice how all these numerals have the Arabic feminine form, which is the normal way for using them in it also.

4.8 The Numerals Eight and Nine

The English numeral *eight* (German *acht*, Latin *octo*, Greek *okto*, Sanskrit *astau*) does not seem to have a phonetically and semantically obvious cognate in Arabic, which is

thaman(iat). As can be seen, both words have no similarity in form to be linked together. However, on the basis of form, *eight* is directly derived from the next higher Arabic numeral *tis3(at)* (nine) or *tasi3* (ninth) in which sound and semantic changes occurred. First, the sounds of *tis3(at)* had been completely reversed or reordered in which /3/ and /s/ merged as /gh/ or /k/ whereas initial and final /t/, an optional variant, collapsed. Next, lexical shift led the numerals *eight* and *nine* to swap reference with each other. That is, *nine* in Arabic is *eight* in English and vice versa. A schematic representation of the phonetic course of evolution for *tis3(at)* in English might look like *tis3a → 3ast → ast → asht → act → eight*.

4.9 The Numerals Nine and Eight

As with the numeral *eight* above, *nine* seems to have no obviously direct Arabic cognate. However, *nine* (Latin *novem*, French *neuf*, Greek *ennea*) and *eight* have exchanged their reference. In other words, Arabic *eight* is English *nine* and vice versa. Therefore, *nine* is a direct and straightforward cognate of Arabic *thaman(iat)* (eight) and related derivatives like *thamin, thumn* (eigth, 1/8), *muthamman* (eight-sided). Phonetic change caused the merger of /th/ and /m/ into /n/, yielding *nine* in English. The course of development for *nine* in English might look like this: *thamani → famani → mani → nani (nine)*.

As regards the Latin and Greek forms, the former shows reordering in which /th/ became /v/ whereas the latter exhibits the merger of /th, m, & n/ into /n/. The French form is like the Latin one where /th/ became /f/. Again in all cases, semantic swap or lexical shift took place.

4.10 The Numeral Ten

Again there seems to be no obvious one-to-one correspondence between the English and Arabic words here: *ten* versus *3ashra(t)* (ten). However, in view of its relationship to the Greek and Latin cognates *deka* and *decem*, both of which have something to do with the hands and fingers which were and are still used for counting by countless people worldwide, producing such items as *digit, digital, dexterity, index*, etc. (e.g., Harper 2012), then its immediate Arabic cognate is a reversed form of *yad, yadan* (pl.) (hand(s) or 10 fingers) where /d/ turned into /t/ (while /y/ turned into /k/ or /s/ in Greek and Latin). In addition, Arabic has another synonym which is *zind, zunood* (pl.) (hand, arm, a unit of measurement, a rod) in which reversal occurred coupled with the passage of /z/ into /s/ in Latin or /k/ in Greek, /n/ into /m/ in Latin and loss in Greek. Thus, it seems that *zind* is an equally likely cognate. Diagrammatically, the course of the sound change for *ten* in English can be represented as *yad(an) → dayan → dajan/dakan → dahan → tahan → tan* (ten).

Concerning the normal Arabic numeral *3ashr(at)* (ten), its direct and true cognate in English with the same form and meaning is *score* as in *scores* (i.e., tens) *of men* in which /3/ changed to /s/ and /sh/ to /k/.

4.11 The Numerals Eleven-Nineteen

Except for *eleven* and *twelve*, all these numerals are obviously combinations of the cardinals *three* to *nine* and a lengthened form of *-teen* (ten), whose Arabic origins have already been settled (4.3-10). The problem is what to do with the exceptions. According to (Harper 2012),

eleven and *twelve* are compounds of a reduced form of *one* and *two* plus *lif(an)* in Old English, whose meaning is *one left* and *two left*. In Arabic, *fal* (leave) is a cognate to *lif(an)*, though in reverse. However, the proper solution would be to treat them in the same way as the other numerals: i.e., as combinations of cardinal *one* or *two* and *ten*. How?

Eleven (*elf* in German) is a combination of *ele-* (one) plus *–ven* (ten). The first element comes from and is a reduced form of the ordinal Arabic numeral *awwal* (one, first) in which /aw/ shifted or changed to /e/, thus yielding *ele-*. This analysis agrees with and further confirms the argument in 4.1 above that the underlying or original form for *one* is *awwal*. The second element *–ven* is a further mutated pronunciation of *ten* in which /t/ became /v/, whose Arabic cognate has already been settled. In German, *-ven* (zehn) was further reduced to /f/ as a result of merger between /v & n/. Briefly put, the course of sound change for *eleven* in English might be expressed thus: (*awwal* (first, one) → *wal* → *el-*) + (*ten* → *(then)* → *ven*) → *eleven*.

As to *twelve*, it behaves similarly where it consists of *two* plus *–ven* (ten) in which /n/ turned into pre-final /l/. Again, as has been stated in 4.2 above, *two* derives from a reduced form of *ithnan* or *thani* in Arabic.

4.12 The Numerals Twenty-Ninety

The numerals *twenty* to *ninety* are made up by the addition of the suffix *–ty* to the relevant cardinal with slight phonetic or orthographic modifications such as *four → forty*, *three → thirty*. It developed from a further shortening of *-tig* (*-zig* in German) (ten) in which /g/ passed into /y/. As shown in 4.10 above, its Arabic cognate is *yad* (hand) in which reversal occurred as well as the passage of /y & d/ into /g & t/ respectively. *Twenty*, however, seems to diverge a little by inserting /n/, which is actually retrieved from *twin* (*zwēne* in Old High German), thus confirming the analysis in 4.2 above that *two* derives from *twin* and, consequently, from *ithn(at)an* (or *thani*) in Arabic.

Now the question to ask is where the ordinal number suffixes *–th* of *fourth, fortieth,* etc. and *–d* of *third* and *second* came from. Actually, both evolved from the Arabic feminine singular suffix */-at/* on morphological and phonetic grounds. Morphologically, the Arabic numerals 1-10 have masculine and feminine forms as in *thalath* (three- masc.) versus *thalath<u>at</u>* (three-fem.). In other words, the ordinal English suffix */-th/* stemmed from the Arabic feminine singular suffix */-at/*. Phonetically, the phoneme /t/, a voiceless alveolar stop, underwent two changes: one to /th/, a voiceless interdental fricative, and one to /d/, a voiced alveolar stop, with the first type being commoner, though. Thus, */th/* in *seven<u>th</u>* is */-at/* in *sab3<u>at</u>* (seven-fem.), for instance.

4.13 The High Numerals Hundred, Thousand, and Million

The numerals *hundred, thousand* and *million* can all be traced back to two Arabic synonymous numeral words and their derivatives for *hundred*: namely, *hindeed* (or *hind* and *hunaidat*), an obsolete word, and *miat* (pronounced *miyat*), the common and usual one. Now *hundred* is directly derived from the classical Arabic word *hindeed* (of camels, a hundred) in which /r/ is an insertion or a split from /n/. The well-known *kentum-centum* or *centum-satem* divide *(*e.g., *cent, century, centenary, centennial, percentage, percentile)* in European

languages which are subdivided into Western (e.g., Italic, Germanic, Hellenic) and Eastern (e.g., Slavic, Indic, Iranian) branches (Campbell 2006: 157, 162; Crowley 1997: 228-229) can be explained in the same way in which /h/ turned into /k/ or /s/ while /d/ into /t/. Notice how Arabic has two old forms for 100, which can accommodate and account for both *hundred* and *cent*: i.e., one from *hindeed* and one from *hind*. To sum it up formulaically, the above sound changes might have proceeded as (i) *hindeed* → *hindreed* (hundred) for English or (ii) *hind* → *hint* → *sint* (cent) (→ *sit* (sat)) for the others.

As for the English and Arabic numerals *thousand* and *alf*, both do not have any formal similarity to each other. However, *thousand* derives from another two Arabic sources, based on its two meanings (Harper 2012). First, as a compound, it may be analyzed as the sum of ten hundreds: i.e., *thou* (literally, hand; ten) and *sand*, being a mutated form of *cent* where /t/ became /d/. In other words, it derives from the same Arabic cognate for *hundred* plus a reversed form of Arabic *yad* (hand) in which /d/ passed into /th/. Secondly, as a whole word, it means *several hundred* or *great multitude*. Arabic has a formally and semantically similar word which may be its origin: viz., *sawad*, pl. *sudan* (blackness, multitude, majority) as in *sawad al-qaum* (the majority of people) and *sawad katheer* (lit., black much: i.e., many people). Sound changes caused initial /s/ to split into /s/ and /th/ whereas /n/ was either an insertion or a retrieval from the plural form. In short, Arabic *hind(eed)* split into two words, yielding *hundred (*and *cent*) and *thousand*. To express it as a formula, (*yad* (hand) → *day* → *thay*) → thou + (*hind* (hundred) → *sind* (-sand) → *thousand*.

Mile, mille and *million* (plural of *mille* (Harper 2012)) are all related to one another, on the one hand, and to   Arabic *mia(t)* (hundred) and its different plurals *miaat* (also pronounced *miyat*) and *mieen*, on the other, in which /t/ passed into /l/. More precisely, *mile* and *mille* came from the singular form *miya(t)* in which /t/ turned into /l/ whereas *million* from the plural form *mieen* in which /l/ evolved from /t/, split from /n/, or was inserted. Notice how *million* is the plural of *mille* in Latin as *mieen* is the plural of *miat* in Arabic. It is worth noting here that many a good native English speaker, for example, pronounce *million* without /l/ as /miyen/ just as happens in French in similar cases. Furthermore, the Arabic word *meel* (a distance measurement unit) is another likely source for *mile*. In addition, the Arabic word *miat* is directly related to the general numerical word *myriad* (countless, ten thousand) in which /t/ passed into /d/ while /r/ was inserted. In all these words, semantic change or lexical shift happened. The above sound changes can be summed up thus: (i) *miat* (hundred) → *mial* → *mile, mille*, (ii) *mieen* (hundreds) → *million*, and (iii) *miaat* (hundreds) → *miaad* → *miriaad* (myriad).

As to the top numeral words *billion* and *trillion*, they can be expounded in the same manner. The former is a compound from the second syllable of *million* plus the prefix *bi-* (from *binary*), which derives from Arabic *thani* (second, two) or a related derivative in which /th/ became /b/ while /n/ split into /n & r/. Similarly, the latter is made up of the second syllable of *million* plus the prefix *tri-* (three), derived from *thalath* in Arabic as shown in 4.3 above. To these can be added *zillion*, which is a new coinage with an indefinite numerical meaning (Harper 2012), consisting of /z/, the last Roman alphabet, and the second syllable of *million*.

Now what happened to Arabic *alf* (thousand) in English above, one might curiously ask? Well, it is used in the numerical expression *five-fold*, *ten-fold*, etc. How? This word has two meanings in English, one as *times* and one as *tie*, *bend*. Theses two senses cannot be related without considering their Arabic origins. That is, both meanings come from two formally related but semantically different Arabic source roots, which are *alf* (combine, thousand) and *laf, laffat* (n) (fold, tie). The former gave rise to *times* via lexical shift and the latter to *fold, tie* via reordering and the passage of /t/ into /d/.

In summary, all the low and high numeral words in English and other European languages have been successfully traced back to their genetic Arabic origins. Thus all have true Arabic cognates or sources.

5. Discussion and Conclusion

The above description and analysis of all Arabic and English numerals have shown that they are real and true cognates in the sense of being genetically related formally and semantically. However, the basic numerals are 1-10 with all their derivatives being simply modifications or combinations thereof. Thus once relating these to their Arabic counterparts, origins or cognates has been done, everything else becomes so easy and straightforward, indeed. More precisely, three patterns of correspondence between Arabic and English numeral words were found, which were as follows:

a) In *three, six, seven* and *hundred* (as well as *million*), there is a straight correspondence and obvious relationship between form and meaning.

b) In *one, two, four,* and *five*, the correspondence is a little less obvious.

c) In *eight, nine, ten*, and *thousand*, the correspondence was not clear at all.

To dismantle ambiguities and establish the genetic relationships amongst such numerals, phonetic, morphological and semantic factors and processes had to be brought into play. That is, the apparently different forms of the Arabic numerals in English, German and French as well as Latin and Greek happened as a result of linguistic changes at the phonetic, morphological and lexical levels. Phonetically, Arabic sounds were subject in such languages to various modifications and developments such as deletion, addition, reversal, reordering, merger, split, duplication, and so on. The main sound changes that affected Arabic consonants and vowels can be summed up as follows:

(a) the pharyngeal consonants /3/ and /2/ were either deleted, merged with similar consonants or turned into others such as /k/ and /s/ as in the words for *four* and *one* in English;

(b) the glottal stop /ʔ/ was either deleted, merged with similar consonants or turned into others such as /h or w/ as in the words for *one*, *two* and *mile* (*mille*, *million*) in English;

(c) the velar /kh/ changed to /k, s or f/ as in the word for *five* in English and French;

(d) the uvular stop /q/ passed into /k, t, or ch/ as in the word for *four* in Latin, French, Greek and Russian;

(e) the interdental /th/ became /t, d, or s/ as in the words for *three/third* in English, German and French;

(f) the emphatic consonant /T/ changed to /t/ or /d/, its plain counterpart, as in the word for *four* in Latin, French, Greek and Russian;

(g) /l/ turned into /r/ as in the words for *three/third* in English, German and French;

(h) /m/ developed into /n/ or /f/ as in the word for *five* in English and German;

(i) /s/ became /k/ as in the word for *five* in French and Latin;

(j) /h/ changed to /k/ or /s/ as in the word for *cent* in English, French and Latin;

(k) /d/ passed into /t/ as in the word for *cent* in English, French and Latin;

(l) /t/ evolved into /d/ as in the word for *second* in English and Latin;

(m) /r/ was inserted as in *myriad* in English;

(n) the three basic long vowels /a:/, i:/, and /u:/, their short counterparts /a/, /i/, and /u/, and the diphthongs /ay/ and /aw/ all underwent different sound changes by exchanging values amongst one another, including fronting (e.g., /u/ → /i/), backing (e.g., /i/ → /u/), raising (e.g., /a/ → /i/), lowering (e.g., /i/ → /a/), centering (e.g., /i/ → /ə/), lengthening (e.g., /i/ → /i:/), shortening (e.g., /i:/ → /i/), diphthongization (e.g., /i:/ → /ay/) and monophthongization or smoothing (e.g., /ay/ → /i:/). In fact, vocalic changes are simpler than the consonantal ones, which are the primary focus of this research.

In short, the different forms of Arabic numerals in modern European languages such as *three* in English, *trei* in German, *trois* in French or ancient ones like Latin, Greek and Sanskrit, are due to different courses of sound change in these languages. A very long historical development, huge geographical isolation, complete lack of social contact and linguistic interchange aggravated all such differences.

Morphologically, the differences were of two types. First, Arabic has more forms for each numeral, depending on number, gender and case whereas English, German or French have less. This entails that English, for instance, is selective where, for a particular number, it sometimes uses the masculine form, sometimes the feminine form, sometimes the plural form. Secondly, the differences were mainly in the use of the Arabic feminine suffix *–at/-ti* v. *–th* in English and the Arabic plural suffix *–een*. For example, *fourth* corresponds to the Arabic feminine form *arba3at* or *rab3at* as opposed to *four*, which comes from the Arabic masculine form *arba3* or *rub3*; *sev<u>en</u>* (or *sieb<u>en</u>* in German) to *sab3<u>een</u>* (seventy) in Arabic.

Finally, on the lexical level, three or four patterns were noticed: lexical stability, lexical shift, lexical split, and lexical selection. Lexical stability means that most numerals in Arabic and English have the same or similar meanings such as 1-7. Lexical shift, a kind of semantic change, is evident in those numerals which exchanged and swapped their reference or meaning as happened in the case of *eight* and *nine* in English which are the other way round in Arabic; other examples include *myriad* (10 000), *mille* (1000) and *million* in English, all of

which came from *miyat* (hundred) and its two plurals *miaat* and *mieen* in Arabic. That is, *million* came from *mieen* while *mille* and *myriad* from *miat* or *miaat*. Lexical split means a word like *hindeed* refers to both *hundred* and *thousand* or *awwal* (first) from which *one* and *eleven* emanated. As to lexical selection, it involves choice from a list of options, so to speak, since Arabic has two or more variants for certain numerals. For example, English and all Western European languages use the ordinal number for *one* in Arabic which uses its cardinal version. This does not imply that the other versions or related derivatives of Arabic numerals are not used; rather they are employed elsewhere such as the use of the different forms of *miat* in Arabic to refer to *mile, mille,* and *million* in English. Another example is the use of *odd* from *wa2ad/a2ad* (one) or *scores* from *3ashrat* (ten), as has been described in 4.1 and 4.10 above.

Thus all the Arabic and English numerals are real cognates in the sense of having similar forms and meanings. In particular, Arabic was found to be the genetic end source of all such numerals. Why? To answer this question properly, a few logical premises need to be considered. First, if for a particular word or form X, a language has three variants ABC while another has one or two thereof, then the language with all the variants is the origin. In other words, the language with more forms is the origin from which the language with less forms is derived. Secondly, if a language uses all or more of the forms or words more productively as nouns, verbs, and so on while another less so, then the productive one is the origin.

Now, applying these premises to the present analysis, the following patterns have been found. To begin with, Arabic numerals have phonetic complexity as opposed to the phonetic simplicity of those in English, German, French or Latin such as the absence of the pharyngeal consonants /2/ and /3/, the uvular consonant /q/, the emphatic consonant /T/ and the velar /kh/ in most European languages. Also Arabic numerals usually have more syllables than those in English in almost all cases. Secondly, Arabic numerals have a wider variety and multiplicity of forms based on number, gender, case, derivation and lexical options or alternatives. This property accounts for and accommodates all the different forms in English, German, French, Russian, Latin, Greek and Sanskrit. For example, *one* in English is *eka* in Sanskrit but *adin* in Russian while *four* in English but *quatre* in French. How can these forms be related to one another? That would be an impossible task to do; however, they can be successfully related when their Arabic origins are considered. That is, *one, eka* and *adin* are treated as different reflexes of their Arabic variants, which are *awwal* (first, one) and *wa2id(an)* or *a2ad(an)* (one) where the former gave rise to the English, French and German versions whereas the latter to the Sanskrit and Russian forms in which /2/ either turned into /k/ or was deleted. In the case of *four* and *quatre*, they belong to two different Arabic cognates which are *arba3* and *qeeraT* as shown in 4.4 above.

Thirdly, Arabic numerals can have a full grammatical, morphological and lexical behaviour where various forms are used, depending on number (singular, plural), gender (masculine, feminine), case (subject, object) and derivation (noun, verb, adjective and adverb). For example, the Arabic numeral *sit* (six) may have the following forms:

- *sit* (six- masc.), *sittat* (six- fem.), pl. *sittaat* & *astaat*;

- *suds* (n), pl. *asdas* (one-sixth);

- *sadis* (adj.), pl. *sawadis* (sixth);

- *sadisan* (adv.) (sixthly);

- *sudasi* (adj.) and *musaddas* (adj.) (six-sided);

- *tasdees* and *tasteet* (n) (classifying into sixes);

- *sattat, saddas* (v) (to make six).

In English, although some numerals may have a limited number of such forms such as *one, oneness, only, alone, unique, unity, unify, unison,* most do not do so such as *four, five, six*, etc. Even then the total number of Arabic forms and derivatives would be far too much larger. For example, if you were to count the different forms for *two* in English and its counterpart *ithn(at)an* in Arabic, you would find that the former has about ten forms at the most (e.g., *two, duo, duet, dual(ity), double, duplicate*) whereas the latter has a lot more.

Finally, Arabic numerals still retain their root-based semantic identity or meaning. In other words, all Arabic numerals have meanings of their own other than expressing the quantity or number *one, two*, etc. On the basis of their roots, they express senses, concepts, ideas which have later been associated with what the current numbers refer to. For example, the Arabic numeral *sab3(at)* (seven) is denotatively and connotatively related to *sab3* (tiger, wolf, courageous, fast), *sab3een* (70, a lot), swearing (e.g., *sab3(ah)u* (Damn him; frightened him); the Arabic numeral *wa2id* (one) derives from a root which means *single, unify, avoid,* and so on; the word *3ashra(t)* (ten) comes from the root *3ashr* which may mean, amongst others, *associate with; to fondle, to play around with; of animals, to get pregnant*, etc.; the word *alf* (thousand) is derived from a root, which means *gathering, combination, familiarity*, etc. This, it seems, is how some numerals at least might have originated in the first place. They are words; they have roots, which finally mean something related. On the other hand, in English, German, French and Greek or Latin numerals, meaning is not as transparent and retrievable. In some numeral words, it is, however. For example, the numerals *quarter, ten,* and *thousand* derive from Latin and Greek words which mean *a cut, a hand* and *a great multitude* in that order, as indicated earlier. All this signifies that numerals were not numerals in essence perhaps, a property they acquired later as these examples indicate. Thus, for all the above reasons, Arabic is the source origin from which English and European numerals descended.

Now the question to ask is whether Arabic and English belong to the same family or language. In other words, are they members of a language family or dialects of a language? To answer this, let's consider Crowley's (1997: 173, 182, 184) classification of languages into sub-groups based on a 100-word list, which is as follows:

Level of subgrouping	% Shared cognate core vocabulary	Years of Separation
Dialects of a language	81-100	-less than 500 Years
Languages of a family	36-81	-500-2500 Yrs

Families of a stock	12-36	-2500-500 Yrs
Stocks of a microphylum	4-12	-5000-7000 Yrs
Microphyla of a mesophylum	1-4	-7000-10000 Yrs
Mesophyla of a macrophylum	0-1	-over 10000 Years

On the basis of the above classification and years of separation concerning the use of numerals, for languages of a family and dialects of a language, the percentage of shared core vocabulary should be between 36-81 and 81-100 each. Now the present analysis has shown that *all* the numerals, whether low or high, are real cognates. Therefore, it can be safely said that in the semantic area of numerals, Arabic and English are dialects of the same language since all the numerals are used in broadly similar forms and meanings. However, this does not mean that their separation is less than 500 years; instead, it is very, very, very much older than even the highest date given for all. When exactly? Only Allah or God on High knows!

But the most important question next is why both languages are not mutually intelligible. Why don't speakers of Arabic, English and French understand each other? That's a very logical and plausible question, indeed. There are three main reasons for that. To start with, phonetic change is cyclic in the sense that more than one sound process was applied to any given numeral, thus making it extremely difficult to recognize. For example, *three*, a rather obvious cognate to *thalath* in Arabic, involved three consonantal sound changes: (i) final /th/-loss or merger with initial /th/, (ii) turning /l/ into /r/, and (iii) turning /th/ into /d/ in *third*. A fourth vocalic change involved turning /a/ into /ee/. In modern Arabic dialects, only one sound change is involved at a time where *thalath* (three) becomes *talat* in Damascus Arabic or *salas* in Egyptian Arabic, for instance (for a survey, see Jassem 1993, 1994a, 1994b). Secondly, the morphological structure of the numerals differs from one language to another. Thirdly, lexical shift led some numerals to swap or change their meanings. To this one can add lexical selection where different words are used for different numerals such as *quatre* in French, *four* in English, and *arba3* in Arabic. Finally, geographical isolation and lack of social contact exacerbated the situation, let alone racial and cultural bias.

In conclusion, one can say with confidence that Arabic is the origin of all European numerals. This has huge theoretical implications as to the validity and practicality of postulating a proto-Indo-European origin for European languages, which is a mirage with no existence in reality. Such a hypothetical origin has entailed putting up not only walls of language separation, segregation and even apartheid but also walls of racial and cultural bias, prejudice and arrogance amongst peoples on both sides of the Mediterranean for hundreds of years (cf. Trudgill 2000: pp. 43-44). What is the use of hypothesizing or founding a superfluous theoretical construct or proto-type when the real source is at a stone's throw away from one? Are not Europe and Arabia or the Arab lands closer to each other than to any other area in the world? Are not their histories more closely enmeshed and intermingled with one another than with anything else? Aren't they geographical and social neighbours? Which is closer to Rome, Paris or London after all: Damascus, Cairo and Baghdad or New Delhi and the Himalayas? Whatever the answer may be, Arab cities are in the middle between East and West. Therefore,

as the languages of these areas are very, very close as has been shown here, such walls and barriers must come down in the interests of knowledge, truth, peace, humanity and unity of the human race. What facilitates that is the fact that Arabic and European languages and peoples are too close geographically, culturally, linguistically and even physically.

Thus, in light of all the above, the lexical root theory has successfully proven to be adequate for analyzing and establishing the genetic relationship between Arabic numerals and their counterparts in English and other European languages where Arabic was found to be their main origin, indeed. To further substantiate that, more research is needed on all language levels, especially the lexical. In fact, the present author has already accomplished a great deal of work in this area phonetically, morphologically, grammatically and lexically, especially the first three. In addition, applications of these findings to language teaching, cultural awareness and understanding are badly needed because they can bring minds and hearts closer than ever before, thus bringing peace, security, stability and harmony to (especially this part of) the world. In short, this is a virgin territory for research which needs to be investigated in depth in all kinds of ways and hues.

Acknowledgement

References

Campbell, L. (2004). *Historical linguistics: An introduction. (*2nd ed). Cambridge, Mass.: The MIT Press. pp. 126, 190-191, 201-202

Crowley, T. (1997). *An Introduction to historical linguistics*. (3rd ed). Oxford: Oxford University Press. pp. 171, 173-174, 182-184

Crystal, D. (2010). *The Cambridge encyclopedia of language*. (3rd ed). Cambridge: Cambridge University Press. p. 308

Harper, Douglas. (2012). *Online etymology dictionary*. [Online] Available: http://www.etymonline.com/index (June 11, 2012)

Ibn Manzoor, Abi AlFadl Al-Misri. (1955-1974). *Lisan al3arab*. Beirut : Dar Sadir.

Ibn Seedah, & Ali bin Ismail. (1996). *AlmukhaSSaS*. Beirut: Daar I2ya Alturath Al3arabi and Muassasat Altareekh al3arabi.

Jassem, Zaidan Ali. (1993). *Dirasa fi 3ilmi allugha al-ijtima3i: Bahth lughawee Sauti ijtima3i fi allahajat al-3arabiyya al-shamiyya muqaranatan ma3a al-ingleeziyya wa ghairiha*. Kuala Lumpur: Pustaka Antara.

Jassem, Zaidan Ali. (1994a). *Impact of the Arab-Israeli wars on language and social change in the Arab world: The case of Syrian Arabic*. Kuala Lumpur: Pustaka Antara.

Jassem, Zaidan Ali. (1994b). *Lectures in English and Arabic sociolinguistics*. Kuala Lumpur: Pustaka Antara.

Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press.

Pyles, T., & J. Algeo. (1993). *The origins and development of the English language. (*4th ed)*. San Diego: Harcourt Brace Jovanovich. pp. 76-77

Ruhlen, M. (1987). *A guide to the world's languages: Classification*. London: Arnold.

Ruhlen, M. (1994). *On the origin of languages: studies in linguistic taxonomy*. Stanford, Ca.: Stanford University Press.

Trudgill, P. (2000). *Sociolinguistics: An introduction to language and society*. (4th ed). London: Penguin Books. pp. 43-44

Viney, B. (2008). *The history of the English language*. Oxford: Oxford University Press.

Copyright Disclaimer