

**NSM substantives:
The Arabic and Hebrew exponents of six simple, universal concepts**

Sandy Habib
Department of English
The Ohalo Academic College, Israel
mhsandy3@gmail.com

ملخص

بحسب نظرية اللغة-المعرفة الدلالية الطبيعية، فقط ستة من أصل كل الأسماء والضمائر تُعدُّ أسماءً وضمائر بسيطةً وعالميةً. في اللغة الإنكليزية، هذه الأسماء والضمائر الستة هي *I, you, someone, something, people, body*. البرهان على بساطتهم ينبع من عدم القدرة على تعريفهم بواسطة كلمات أبسط. أما إثبات عالميتهم فيحتاج إلى إيجاد نظرائهم في أكبر عدد ممكن من اللغات. تهدف هذه المقالة إلى فعل ذلك في ثلاث لغات سامية، وهي اللغة العربية الجشية، اللغة العربية الفصحى، واللغة العبرية، فنثبت بذلك أن لهذه الأسماء والضمائر الستة ذات النظراء في اللغات الثلاث.

Abstract

Of all the substantives, only six are regarded by the Natural Semantic Metalanguage (NSM) as being simple and universal. These six substantives are realized in English by means of the words *I, you, someone, something, people, and body*. While proving their simplicity is evidenced by the fact that they cannot be defined further using simpler terms, proving their universality requires identifying them in as many languages as possible. This paper aims to do so in three Semitic languages, which are Jish Arabic, Standard Arabic, and Hebrew, and demonstrates that these six concepts indeed have exponents in these three languages.

Keywords: Natural Semantic Metalanguage; substantives; Jish Arabic; Standard Arabic; Hebrew.

1. Introduction

1.1 The Natural Semantic Metalanguage

The Natural Semantic Metalanguage (NSM) is a semantic theory developed by Anna Wierzbicka, Cliff Goddard, and colleagues over a period of about five decades. The most fundamental concept in NSM is that of “semantic primes.” It refers to meanings which are simple and indivisible, namely, to meanings which cannot be defined in simpler terms. Consider, for example, English *see*, as in the sentence ‘I can see you now.’ This English word is regarded as an exponent¹ of a semantic prime, and it seems that this word resists decomposition into simpler words. To clarify this claim, I will present three dictionary definitions of this word and demonstrate how they fail to decompose it using simpler words.

The *Cambridge Advanced Learner’s Dictionary* (2005: 1147) defines *see* (in the relevant sense) as “to be aware of what is around you by using your eyes.” The *Longman Dictionary of Contemporary English* (2003: 1483) defines it as “to notice or examine someone or something, using your eyes”. The *Merriam-Webster Dictionary* (2003: 1303) defines it as “a. to perceive by the eye; b. to perceive or detect as if by sight.” As can be seen, the three dictionaries use words which are more complex than *see* (such as, *aware*, *notice*, and *perceive*) to define this word, which results in obscurity. In addition, the three dictionaries use the word *eye* or *sight* in the definitions. When consulting these dictionaries regarding the meanings of these two words, we find that they are defined via *see*, which results in circularity. Thus it is not hard to see that defining *see* using simpler words appears to be unfeasible.

Semantic primes have been identified in typologically and genetically different languages (Goddard and Wierzbicka 1994; Goddard and Wierzbicka 2002a; Goddard 2008; Peeters 2006; Levisen 2013; Gladkova 2010), thus underpinning the premise that they are universal human concepts. The rationale behind using semantic primes as tools of semantic analysis is that simple words have to be used when defining a certain word; otherwise, circularity and/or ambiguity are likely to creep into the definition of the word in question. Moreover, many concepts in a given language are not found in other languages; to explain these concepts to cultural outsiders or learners, universal words have to be used. Semantic primes, being simple and universal, can help in escaping circularity and ambiguity and in constructing definitions that are translatable into any language. The number of these primes currently stands at 65, as can be seen from the table below:

¹ An ‘exponent’ of a semantic prime in a certain language is the realization or lexicalization of that semantic prime in that language.

Table 1: English exponents of semantic primes (Goddard and Wierzbicka 2014)

English exponents	Semantic group
I, YOU, SOMEONE, SOMETHING~THING, PEOPLE, BODY	Substantives
KIND, PART	Relational substantives
THIS, THE SAME, OTHER~ELSE	Determiners
ONE, TWO, MUCH~MANY, LITTLE~FEW, SOME, ALL	Quantifiers
GOOD, BAD	Evaluators
BIG, SMALL	Descriptors
THINK, KNOW, WANT, DON'T WANT, FEEL, SEE, HEAR	Mental predicates
SAY, WORDS, TRUE	Speech
DO, HAPPEN, MOVE, TOUCH	Actions, events, movement, contact
BE (SOMEWHERE), THERE IS, BE (SOMEONE/SOMETHING), BE (SOMEONE)'S	Location, existence, possession, specification
LIVE, DIE	Life and death
WHEN~TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT	Time
WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE	Space
NOT, MAYBE, CAN, BECAUSE, IF	Logical concepts
VERY, MORE	Intensifier, augmentor
LIKE	Similarity

Different exponents of the same semantic prime—in any given natural language—are called *allolexes*. There are two main kinds of allolexes, viz. positional and combinatorial. A positional allolex is an allolex that is called for when a different exponent form is needed because of the changing of the position of the semantic prime in the sentence, e.g. nominative I vs. accusative or oblique ME. A combinatorial allolex is an allolex that is needed because of the combination of the semantic prime with certain words; for instance, the English exponent THIS is

used with singular nouns while the allolex *THESE* is needed to modify plural nouns (Goddard and Peeters 2006).

As has been mentioned, semantic primes have been investigated and identified in a number of languages. Yet, they have not been identified in Arabic and Hebrew, two major Semitic languages. Investigating all the Arabic and Hebrew semantic primes in one paper is impossible owing to space constraints. Therefore, in this paper, I will explore the exponents of the substantives (I, you,² someone, something, people, and body) in Standard Arabic, Jish Arabic, and Hebrew. Identifying these (as well as the other) exponents, not only enables the translation of explications into Arabic and Hebrew, but it also lends support to the premise that semantic primes are universal.

1.2 Standard Arabic, Jish Arabic, and Hebrew

Arabic is the official language of 22 Arab countries as well as one of the official languages of 8 other Arab and non-Arab countries, and it is spoken by about 422 million people (Holes 2004).³ Arabic is diglossic, and has two genetically-related varieties, standard and non-standard (cf. Schiffman 1997). In contrast to its non-standard counterpart, the standard variety consists of one form, and this form is the one used in written literature, news, education, and most formal settings (e.g. courtrooms). Also, unlike the non-standard variety, the standard variety is not spoken natively by anyone. This raises a problem in regard to “naturalness.” If a language is used by a community of people in certain domains/settings, but it is not spoken natively by any member of this community, how can members of this community decide whether a certain utterance in that language is natural or not, especially if it falls outside the domains in which that language is used?

The non-standard variety comprises a myriad of forms, and each form is spoken natively by a certain community. Non-standard forms which are geographically adjacent to each other are mutually intelligible. In theory, the larger the geographical distance between two non-standard forms, the less intelligible one form is to the speakers of the other form. However, owing to globalization and media, certain forms, such as those of Cairo and Beirut, seem to be highly comprehensible to all adult Arabs (Watson 2002; Palmer 2007). In this paper, I will investigate, besides Standard Arabic and Hebrew, a non-standard form of Arabic which I, among other

² The second-person singular pronoun (realized as *thou* in earlier forms of English).

³ See also: “List of Countries Where Arabic is an Official Language”. Retrieved from http://en.wikipedia.org/wiki/List_of_countries_where_Arabic_is_an_official_language (accessed October 11, 2018).

people from Jish, speak natively. Jish is a small town that lies in the north of Israel, about 10 kilometers from the Israeli-Lebanese borders, and has a population of about 3,000 people. Akin to all non-standard forms, Jish Arabic is not officially recognized as a language, and therefore, is not used in writing, except in text messaging between friends, some local commercials, and the like. Jish Arabic does not have a writing system of its own (cf. Habib and Kurzon 2008).

Hebrew is one of the official languages of Israel, and it is the mother tongue of about six million Jewish Israelis. What I call ‘Hebrew’ in this article is called ‘Modern Hebrew,’ ‘Israeli Hebrew,’ or ‘Israeli’ by others (Rozen 1962; Junger 1987; Zuckermann 2009). Hebrew is a relatively modern language which was revived in the late 19th century by Eliezer Ben Yehuda (1858-1922), but it was not until the beginning of the 20th century that it was first spoken natively by a group of Jews residing in present-day Israel (Kutscher 1982; Saenz-Badillos and Elwolde 1993). While the roots of Hebrew are found in Biblical Hebrew, some, if not many, of its semantic and syntactic features are different from those of Biblical Hebrew. This can be ascribed to the fact that Modern Hebrew is based not only on Biblical Hebrew but also on Yiddish, with both being considered primary contributors to Modern Hebrew (Zuckermann 2004; Zuckermann 2009; Zuckermann 2005).

2. Methodology

Before starting to identify the exponents of the substantives, I should draw attention to the method that I use to identify these exponents. The method I employ relies on the following guidelines:

1. In a given language, if (a) there appears to be only one candidate for a certain semantic prime, and (b) this candidate satisfies all the combinatorial frames⁴ of the semantic prime, then this candidate is to be regarded as the exponent of this semantic prime.
2. If (a) there appear to be two or more candidates for a certain semantic prime, but (b) only one of them satisfies all the combinatorial frames of the semantic prime, then only this candidate is to be regarded as the exponent of this semantic prime.
3. If (a) there appear to be two or more candidates for a certain semantic prime, and (b) all of them satisfy all the combinatorial frames of the semantic prime, but (c) only one of

⁴A combinatorial frame is a sentence that is simple and universal. It cannot be simplified further, and it has exact equivalents in all languages.

them seems to be semantically simpler than all the others, then only this candidate is to be regarded as the exponent of the semantic prime.

4. If (a) there appear to be two or more candidates for a certain semantic prime, and (b) all of them satisfy all the combinatorial frames of the semantic prime, and (c) there does not appear to be any semantic difference between them, but (d) one of them is much more frequent than all the others, then only this candidate is to be regarded as the exponent of the semantic prime.⁵

5. If (a) there appear to be two or more candidates for a certain semantic prime, and (b) all of them satisfy all the combinatorial frames of the semantic prime, and (c) there does not appear to be any semantic difference between them, and (d) they are not (much) different in their frequency, then all of them are to be regarded as alloloxes of the exponent of the semantic prime.

Guidelines (1)-(3) were used in the identification of exponents of semantic primes in the languages that had been studied so far. Guidelines (4) and (5) have not been used before. The rationale behind using frequency to rule out candidates stems from the notion that infrequent words may not be known to young children, and this, I argue, may indicate that they are not (conceptually) simple. It can also be the case that neither of the candidates is more complex than the others, but, because one is at least twice more frequent than each of the others, I select that candidate as the exponent of the semantic prime for matters of convenience. Let us assume that, in a given language, we have found 10 candidates for a certain semantic prime, and the only difference between them is that one of them is at least twice more frequent than every other candidate. Positing all of them as alloloxes of the same exponent of that semantic prime might not be practical, because, in the end, we just need to use only one of them in explicating other words. If so, there is no benefit in positing all the others as alloloxes of the same exponent.

To check frequency, I will employ corpus analysis. A corpus is a huge body of texts coming from different sources. To check frequency in Standard Arabic, I will rely on the ArabiCorpus (150+ million words), and, to check frequency in Hebrew, I will use HebrewCorpus

⁵ The phrase ‘much more frequent’ is certainly vague: what number of occurrences would determine the difference between ‘not much more frequent,’ ‘more frequent,’ and ‘much more frequent’? The answer is in no way simple. In this paper, I abide by the following: Of two candidates, if the first is at least twice more frequent than the other, I regard the first as the exponent. I should note that I will not always be able to check the frequencies of the candidates. This is because Arabic and Hebrew have consonantal writing systems, and therefore, a given word can highly likely be homonymous. Checking the frequency then would require a meticulous checking of every context, a task that can be tremendously time-consuming.

(170+ million words).⁶ In addition, I will use Google. The use of Google is justifiable on the grounds that the corpora that I use are relatively small in terms of their vocabulary size. Google, which gives access to a massive number of texts, can, in certain cases, fill the gaps in the corpora. Yet, for a searched word (or clause), Google does not present the numbers of occurrences of this word. Rather, it presents the number of URLs that include this word. This number is called ‘hit counts’ or ‘hits’ (Fletcher 2007; Fletcher 2004). The problem with presenting the number of URLs rather than the number of occurrences is that one URL can have more than one occurrence of this word. Thus the number of hits is not tantamount to the number of occurrences of the word/clause searched for. One way is to conduct several checks of the same word/clause at different times (Goddard 2009; Wierzbicka 2009; Wierzbicka 2010).

In the next section, I attempt to provide a relatively detailed account of the Arabic and Hebrew exponents of NSM substantives. Owing to space constraints, however, I do not always deal with all the complexities that arise. For the same reason, I also do not provide all the possible combinatorial frames and allolexes for each exponent.

3. Substantives

In linguistics, a substantive “is a term used in the classification of words, referring to the class of nouns (traditionally defined as ‘substances’, i.e. names of persons, places, things, etc.), and also to those items which function as nouns, though lacking some of the formal characteristics of that class (cf. the ‘substantival function’ of adjectives, in the poor, the rich, etc.). The set of pronouns may also be included in this class” (Crystal 2008: 463). In NSM, the substantives are six in number, and they are: *I*, *you*, *someone*, *something*, *people*, and *body*. In this section, I shall identify the exponents of these semantic primes in Standard Arabic, Jish Arabic, and Hebrew.

3.1 I and YOU

3.1.1 Standard Arabic: ANĀ and ANTA. The basic exponent of I in Standard Arabic is ANĀ. Because Arabic is a pro-drop language, this independent pronoun is not frequently used. When it is used, however, it occurs mostly in subject position, as in example (1). In all other positions, viz. accusative and genitive, the clitic =I, which is a positional allolex, is used (see

⁶ It is worth noting that these corpora are constantly updating.

example (2)). Besides this positional allolex, there are a number of obligatory portmanteaus that appear as verbal inflections marking subject-verb agreement. In the perfective paradigm, the suffix -TU is attached to the verb; in the imperfective paradigm, the prefixes A- (or U-) is used.⁷

- (1) *anā ra'ay-tu shay'-a-n.*
 I see-1SG something-ACC-INDF
 'I saw something.'
- (2) *anta qul-ta l=ī shay'-a-n.*
 you(2SG.M) say-2SG.M to=GEN.1SG something-ACC-INDF
 'You said something to me.'

In the second person, there is a gender distinction. The masculine form is ANTA and the feminine form is ANTI; both of these forms are used in subject position, as in example (3). In all other positions, such as in (4), the clitics =KA (masculine form) and =KI (feminine form) are used. In addition to these two positional allolex forms, there are a number of obligatory portmanteaus that are used to mark subject-verb agreement. In the perfective and imperfective paradigms, the following affixes are used, respectively (feminine forms appear in parentheses): -TA (-TI) and TA-(TA-...-ĪNA).

- (3) *anta ra'ay-ta shay'-a-n.*
 you(2SG.M) see-2SG something-ACC-INDF
 'You saw something.'
- (4) *anā qul-tu la=ki shay'-a-n.*
 I say-1SG to=GEN.2SG.F something-ACC-INDF
 'I said something to you.'

The gender distinction may seem to result in a problem: which form, the masculine or the feminine, is simpler? I would argue that neither form is more complex than, and thus definable via, the other. One cannot define the masculine form using the feminine form, or vice versa. One cannot also define both forms using such words as *boy* and *man* for the masculine form, as well as *girl* and *woman* for the feminine form. In fact, two problems arise from defining, say, the masculine form ANTA using the feminine form ANTI and words such as *boy* and *man* (as in the definition below). First, not using the negative marker in component (b) results in a contradiction,

⁷ Note that Arabic is usually described as having two morphological tenses, perfective and imperfective (see, e.g., Ryding 2005). The affixes that attach to the verb are aspectual particles that carry agreement only (Aoun et al. 2010)

since *anta* refers to a male person while *anti* to a female person. On the other hand, including the negative marker in component (b) results in ambiguity because, if a person is not you_{feminine}, this person can be you_{masculine}, he, or she. Second, the words *boy* and *man* are themselves more complex than the second person singular pronouns.

ANTA

- (a) someone,
- (b) this someone is (not) *anti*,
- (c) this someone can be a boy [m], this someone can be a man [m]

The two forms *anta* ‘you(SG.M)’ and *anti* ‘you(SG.F)’ do carry information. However, this information can be likened to information carried by masculine and feminine voices, which is not part of the speaker’s intended meaning (cf. Wierzbicka 2002). But how can these two forms be explained to someone (say, an English-speaking child) whose language does not make the same kind of distinction made in Arabic? One possible way is to explain *anta* ‘you(SG.M),’ for example, as “YOU, when someone wants to say something to a man [m], this someone says the word ANTA; this someone says the same word if this someone wants to say something to a boy [m].” The explanation should begin with stating that *anta* ‘you(SG.M)’ means ‘you’; otherwise, the sentences above can be mistaken for explaining, say, the word *mister*.

3.2.2 Jish Arabic: ANĀ and INTI.

Like in Standard Arabic, in Jish Arabic I is expressed as ANĀ and its positional allolex as the clitic =I. Obligatory portmanteaus that mark subject-verb agreement in the perfective and imperfective paradigms are realized as the affixes -ET and BA-, respectively.⁸ The infix <Ā> is used to express the progressive aspect.

- (5) *anā shuf-et ishī.*
I see-1SG something
‘I saw something.’
- (6) *inti ul-ti-ll=ī ishī.*
you(SG) say-2SG.M-to=GEN.1SG something
‘You said something to me.’

⁸ Note that these affixes are but a few of the affixes that are used to mark the different subject-verb agreements in the perfective and imperfective paradigms; space does not permit me to present all these affixes.

The exponent of YOU is INTI. There is no gender distinction when using this pronoun in subject position. In all other positions, however, the masculine and feminine clitic forms, i.e. =AK and =EK respectively, are used as positional allolexes. To convey subject-verb agreement in the perfective and imperfective paradigms, the affixes -ET (-TI) and BTI- (BTI-...-I) are used, respectively (feminine forms are in parentheses). The infix <Ā> is used to express the progressive aspect.

- (7) *inti shuf-et ishī.*
 you(SG) see-2SG.M something
 ‘You saw something.’
- (8) *anā ul-ti-ll=ek ishī.*
 I say-1SG-to=GEN.2SG.F something.
 ‘I said something to you.’

3.2.1.3 Hebrew: ANI and ATA.

In Hebrew, there are two candidates for I, i.e. *ani* and *anokhi*. The latter is less frequent and is found in literary texts only; HebrewCorpus presents 234,703 instances of *ani* but only 7,328 instances of *anokhi*. As a consequence, the basic exponent of I is ANI. The form ANI is used only in subject position. In all other positions, the clitic=I is used. The obligatory portmanteaus -TI, <O>, and E- mark subject-verb agreement in the past, present, and future tenses, respectively.⁹

- (9) *ani ra`i-ti mashehu.*
 I see-PST.1SG something
 ‘I saw something.’
- (10) *ata amar-ta l=i mashehu.*
 you(SG.M) say-PST.2SG.M to=GEN.1SG something
 ‘You said something to me.’

The exponent of YOU is ATA, which is the masculine; the feminine form is AT. As is the case with Standard Arabic ANTA [you(SG.M)] and ANTI [you(SG.F)], there is no reason for claiming that Hebrew ATA [you(SG.M)] is more basic than AT [you(SG.F)], or vice versa. These two forms are used in subject position. In all other positions, the forms =KHA and =KH are used

⁹ Note that these affixes are but a few of the affixes that are used to mark the different tenses; space does not permit me to present all these affixes.

for males and females, respectively. To mark subject-verb agreement in the past, present, and future, the following obligatory portmanteaus are used respectively (feminine forms are in parentheses): -TA (-T), <O> (<O>...-A/E(T)), and TE-(TE-...-I).

- (11) *ata ra`i-ta mashehu.*
 you(2SG. M) see-PST.2SG.M something
 ‘You saw something.’
- (12) *ani amar-ti la=kh mashehu.*
 I say-PST.1SG to=GEN.2SG.F something
 ‘I said something to you.’

3.2.2 SOMEONE and SOMETHING

SOMEONE and SOMETHING are two semantic primes that are needed in the explication of any living being or non-living thing. This is because, in NSM, a living being or a non-living thing can be categorized as either someone (such as human beings) or something (such as animals and inanimate things). In what follows I will identify the exponents of SOMEONE and SOMETHING in Arabic and Hebrew.

3.2.2.1 Standard Arabic: AḤAD and SHAY’.

SOMEONE and SOMETHING are expressed in Standard Arabic by AḤAD and SHAY’, respectively. AḤAD has ĀḤĀD as its plural form, though this form is rarely used. The plural form of SHAY’, on the other hand, is ASHYĀ’, and it is very frequently used. Thus the Arabic exponents seem to behave almost like their English counterparts.

As for combining with specifiers, SHAY’ can occur with both determiners and quantifiers perfectly well. When it occurs with specifiers or in negative constructions, SHAY’ does not have any combinatorial allolexes, as can be seen from example (13).

- (13) *lam a-sma’ shay’-a-n.*
 NEG.PST 1SG-hear something-ACC-INDF
 ‘I didn’t hear anything.’

AḤAD, on the other hand, cannot occur with quantifiers (except for ONE) because it does not have a plural form. The combination of AḤAD with determiners not only sounds unnatural but also misleading. This is because determiners in Standard Arabic require that the nouns they

modify take the definite article *al-* ‘the.’ When the word AḤAD is combined with the definite article *al-*, its meaning changes into the Arabic near equivalent of the English word *Sunday*. For example, in a sentence such as the one in (14), I doubt whether the reader could understand something other than ‘This Sunday is a bad Sunday.’ Even so, when reading the word AL’AḤAD ‘the someone’ in a larger context, it is hoped that the reader would be able to comprehend its meaning.

- (14) *hādhā l-aḥad-u huwa aḥad-u-n*
 this(SG.M) DEF-someone-NOM 3SG.M someone-NOM-INDF
sayyi’-u-n.
 bad-NOM-INDF
 ‘This someone is someone bad.’

Wierzbicka (1996) states that, in many languages, SOMEONE does not combine readily with the determiner THIS, but she does not deem the combination ‘this someone’ incoherent from a semantic point of view. She explains that the meaning of ‘this someone’ is encoded in a special portmanteau which, in English for instance, is realized as the third person singular pronouns *he* and *she*. Because people are accustomed to using these special portmanteaus, they find the combination ‘this someone’ unnatural. It is worth noting, however, that ‘this someone’ is used in the English language, albeit not frequently; The Corpus of Contemporary American English (COCA; (Davies 2008)), for example, shows 49 occurrences of this phrase. In addition, native English speakers can be easily habituated to this phrase, and hence its use in NSM explications.

Like their English counterparts, both AḤAD and SHAY’ can occur with adjectives, as is shown in example [15]:

- (15) *aḥad-u-n jayyid-u-n ra’ā*
 someone-NOM-INDF good(SG.M)-NOM-INDF see(3SG.M)
hādhīhi l-ashyā’-a ṣ-ṣaghīr-at-a.
 this(SG.F) DEF-something(PL)-ACC DEF-small-SG.F-ACC
 ‘Someone good saw these small things.’

3.2.2.2 Jish Arabic: ḤADĀ and ISHĪ.

SOMEONE and SOMETHING are realized in Jish Arabic as ḤADĀ and ISHĪ. Like their Standard Arabic counterparts, ḤADĀ does not have a plural form whereas the plural form of ISHĪ

is ISHYĀ'. Like Standard Arabic, Jish Arabic does not have a word that can replace ḤADĀ in plural constructions.

ISHĪ can be modified by determiners and quantifiers without any problem. On the other hand, the combination of ḤADĀ with determiners sounds unnatural, but it is not impossible, and it does not create the same or even a similar problem to that in Standard Arabic. As for occurring with quantifiers, ḤADĀ does not combine with any quantifier, except ONE, for the mere reason that it does not have a plural form. Furthermore, ḤADA and ISHI can occur with adjectives.

- (16) *ḥadā mnīḥ shāf hadōl*
 someone good(SG.M) see(3SG.M) this(PL.M)
il-ishyā' li-zghār.
 DEF-something(PL) DEF-small(PL)
 'Someone good saw these small things.'
- (17) *mā smi' -t=esh ishī.*
 NEG.PST hear-1SG=NEG something
 'I didn't hear anything.'

3.2.2.3 Hebrew: MISHEHU and MASHEHU~DAVAR.

The primary Hebrew exponent of SOMEONE is MISHEHU. This exponent differs from its Arabic counterparts in three respects. First, it is morphologically complex; it consists of *mi* 'who,' *she* 'that,' and *hu* 'he.' Second, it has combinatorial allolexes, which are MISHEHI [someone(SG.F)], MISHEHEM [someone(PL.M)], and MISHEHEN [someone(PL.F)].¹⁰ Third, it has the two allolexes AFEKHAD [no one(SG.M)] and AFAKHAT [no one(SG.F)], which are used in negative constructions.

- (18) *mishehu ra'a et ze.*
 someone see(PST.3SG) PREP this(SG. M)
 'Someone saw this.'

¹⁰MISHEHEM [someone(PL.M)], and MISHEHEN [someone(PL.F)] are rarely used in Hebrew. HebrewCorpus presents 7 contexts in which MISHEHEM [someone(PL.M)] occurs and two contexts in which MISHEHEN [someone(PL.F)] appears. A Google search, however, yielded 10,100 occurrences of MISHEHEM [someone(PL.M)] on Nov. 12, 2018 and 8,500 occurrences on Dec. 5, 2018. It also resulted in 6,820 occurrences of MISHEHEN [someone(PL.F)] on Nov. 12, 2018 and 6,490 occurrences on Dec. 5, 2018. To illustrate the infrequency of these two forms, compare their occurrences in Google with that of MISHEHU [someone(SG.M)], which appeared 5,050,000 times on Nov. 12, 2018 and 5,210,000 on Dec. 5, 2018.

SOMETHING has two candidates, *mashehu* and *davar*. *Mashehu* is monosemous, whereas *davar* is polysemous and has several meanings, among which are *something* and *object*. *Davar* ‘something’ can be used with all the determiners and quantifiers, but *mashehu* cannot. *Mashehu* does not have a plural form, and therefore, cannot be used with any quantifier, save for ONE. It cannot also be used with the determiner THE SAME. Even though *davar* seems to be a better candidate for semantic primehood than *mashehu*, it should be noted that *mashehu* cannot be defined through *davar*, and neither can *davar* be defined via *mashehu*. As a consequence, both should be regarded as allollexes of the exponent.

In the affirmative, *mashehu* can be used without a problem, unless its plural form is required or it occurs after OTO ‘THE SAME,’ as has been mentioned above. Regarding *davar*, it has to be followed by the indefinite article *ma* in the affirmative; otherwise, it might be interpreted in certain contexts as ‘object/thing,’ e.g. *yesh davar betokh ha`aron* ‘There is a *davar* inside the cupboard’. Note that in English (and the same holds for Hebrew) there is a difference between the two sentences: (1) ‘There is something in the cupboard’ and (2) ‘There is an object/a thing in the cupboard.’ The first sentence can imply in some situations that the speaker does not know what is in the cupboard. The second sentence, on the other hand, implies that what is in the cupboard is non-human. Thus, if someone tells another person that ‘there is something in the cupboard’ and then a child comes out from that cupboard, the sentence that the speaker uttered would be valid. But the same does not go for the same situation with the sentence ‘there is an object/a thing in the cupboard.’

In negative sentences, *mashehu* is replaced by its allollex *klum* whereas *davar* is not replaced by any word, but it can be accompanied by the negative word *shum*. As a result, the English sentence “I didn’t feel anything” can be rendered in Hebrew as either (19) or (20):

- (19) *lo hircash-ti klum.*
 NEG feel-PST.1SG anything
- (20) *lo hircash-ti (shum) davar.*
 NEG feel-PST.1SG (NEG) something

3.2.3 PEOPLE

3.2.3.1 Standard Arabic: NĀS.

The English word *people* is usually translated into Standard Arabic as either *nās* or *sha‘b*. The latter, however, is more complex and can be thought of as the near equivalent of the noun

phrase ‘a people,’ which normally refers to people living in the same country and sharing the same religion, beliefs, and/or traditions. *Sha‘b* can be explicated via *nās*, and therefore, the latter is the exponent of the semantic prime PEOPLE:

sha‘b

many people

these people can live in one country [m]

these people can do many things in the same way

these people can think about many things in the same way

NĀS can combine with any determiner and quantifier, except ONE and TWO.¹¹ With these two numerals, the word *shakhṣ* ‘person’ and its plural form *ashkhāṣ* are normally used. However, *shakhṣ*, like its English counterpart *person*, is not an exponent of a semantic prime, because it can be explicated as *aḥadun min naw‘ innāsi* ‘someone of the kind people.’ Another way to say ‘two people’ in Standard Arabic is using the word *ithnān* ‘two.’ This word can mean either the number two or two entities (i.e. two people, animals, etc.).

When NĀS occurs with the determiners THIS and THE SAME or with the quantifiers SOME and ALL, the definite article *al-* is added to it. *Al-* has to be used also when expressing the notion of ‘people in general.’

- (21) *nās-u-n* *kathīr-ūn* *sami‘-ū*
 people-NOM-INDF many-NOM.M hear-3SG.PL.M
 hādhāsh-shay’-a.
 this(3SG.M) DEF-thing-ACC
 ‘Many people heard this thing.’

3.2.3.2 Jish Arabic: NĀS.

The basic exponent of PEOPLE in Jish Arabic is NĀS. Like its Standard Arabic counterpart, NĀS can combine with all the specifiers, except for ONE and TWO. Unlike Standard Arabic, Jish Arabic does not have a near equivalent to Standard Arabic *shakhṣ* ‘person.’ The phrases ‘one person’ and ‘two persons/people,’ for instance, are translated as *wāḥad* ONE and *tnēn* TWO. Furthermore, when the Jish Arabic NĀS occurs with the specifiers THIS, THE SAME, and ALL, the

¹¹ It is worth noting that, while the combination of *two* with *people* is accepted in English, it is not accepted in NSM because the semantic prime PEOPLE refers to people as a collective group rather than individuals (Gladkova 2007).

definite article *el-* has to be affixed to it. Thus the only difference between Standard Arabic and Jish Arabic in this regard is that in the former, but not in the latter, the definite article is attached to NĀS when it occurs with SOME.

- (22) *nās ktār sim‘-ū hādā l-ishī.*
 people many hear-3SG.PL this(3SG.M) DEF-thing
 ‘Many people heard this thing.’

3.2.3.3 Hebrew: ANASHIM.

English *people* is usually translated into Hebrew as *am*, *bneyadam*, or *anashim*. *Am* is the Hebrew near equivalent of Standard Arabic *sha‘band* English *a people*; therefore, it is complex and can be paraphrased via the semantic prime PEOPLE. *Bneyadam* (lit. ‘sons/children of Adam’) is the near equivalent of English *human beings*; it, too, is complex and can be explicated via people: a major component of the explication of *ben adam* ‘a human being’ would be ‘someone of the kind people.’ Having disregarded *bneyadam* and *am*, *anashim* remains the only candidate for primehood.

ANASHIM can occur with any determiner and quantifier, save ONE.

- (23) *harbe anashim sham`-u et ha-davar ha-ze.*
 many people hear-PST.3PL PREP DEF-thing DEF-this(3SG.M)
 ‘Many people heard this thing.’

The fact that ANASHIM can occur with TWO sets it apart not only from Standard Arabic and Jish Arabic, but also from many other languages investigated by NSM researchers (Peeters et al. 2006; Goddard and Wierzbicka 2002b). One reason that can explain this difference is that the Hebrew word *anashim* is the plural form of *ish*, which can refer to either a male human being or a person. Thus, *shney anashim* can mean either ‘two men’ or ‘two persons.’ This stands in contrast to many languages in which the word for PEOPLE has a different stem from the word for *man* or *human being* (Wierzbicka 1996).

One last thing to mention about ANASHIM is that when it combines with THIS and ALL, the definite article *ha-* has to be added.

3.2.4 BODY

The exponents of BODY in Jish Arabic and Hebrew are straightforward, and no allolexy is involved. They are lexicalized as JISIM in Jish Arabic and GUF in Hebrew. In Standard Arabic, there are two candidates for this prime, namely *jism* and *jasad*. The former, however, is to be regarded as the basic exponent of BODY because the latter refers to a human or human-like body and can thus be defined via *jism*. *Jasad* can be defined as *jismu aḥadin* ‘body of someone.’ This is reminiscent of Amharic, where there are two words which translate into English as *body*: *səw nnət* is the exponent of BODY, whereas *gəla* is used to refer to the body of human beings (Amberber 2008).

Examples (24)-(26) below illustrate the use of Standard Arabic JISM, Jish Arabic JISIM, and Hebrew GUF, respectively:

- (24) *hādhā jism=i.*
this(SG.M) body=GEN.1SG
‘This is my body.’
- (25) *hādā jism=i.*
this(SG.M) body=GEN.1SG
‘This is my body.’
- (26) *ze ha-guf shel=i.*
this(SG.M) DEF-body of=GEN.1SG
‘This is my body.’

4. Conclusion

This paper has successfully identified the exponents of six semantic primes in three languages, Jish Arabic, Standard Arabic, and Hebrew. Their identification lends support to the premise that the aforementioned primes are universal. In addition to this, this paper can be considered as a contribution to the field of contrastive semantics, as well as to the study of nouns and pronouns.

References

Amberber, M. (2008). Semantic Primes in Amharic. In C. Goddard (Ed.), *Cross-Linguistic Semantics* (pp. 83–119). Amsterdam: John Benjamins.

- Aoun, E. J., Banmamoun, E., and Choueiri, L. (2010). *The syntax of Arabic*. Cambridge: Cambridge University Press.
- Crystal, D. (2008). *A dictionary of linguistics and phonetics*. 6th Edition. Malden, MA: Blackwell Publishing.
- Davies, M. (2008). *The Corpus of contemporary American English: 520 million words, 1990-Present*. Available online at <http://corpus.byu.edu/coca/>.
- Fletcher, W. (2004). Making the Web More Useful as a Source for Linguistics Corpora. In U. Connor & T. Upton (Eds.), *Corpus linguistics in North America 2002: Selections from the fourth North American symposium of the American association for applied corpus linguistics*, (pp. 191–205). Amsterdam: Rodopi.
- Fletcher, W. (2007). Concordancing the Web: Promise and Problems, Tools and Techniques.” In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web*, (pp. 25–45). Amsterdam: Rodopi.
- Gladkova, A. (2007). *Russian emotions, attitudes and values: Selected topics in cultural semantics*. PhD Dissertation. Canberra: Australian National University.
- Gladkova, A. (2010). *Russian cultural semantics: Emotions, values, attitudes*. Moscow: Languages of Slavonic Cultures.
- Goddard, C. (2008). *Cross-linguistic semantics*. Amsterdam: John Benjamins.
- Goddard, C. (2009). Not Taking Yourself Too Seriously in Australian English: Semantic Explications, Cultural Scripts, Corpus Evidence. *Intercultural Pragmatics*, 6(1), 29–53.
- Goddard, C., and Peeters, B. (2006). The Natural Semantic Metalanguage (NSM) Approach. In B. Peeters (Ed.), *Semantic primes and universal grammar: Empirical evidence from the Romance languages*, (pp.13–38). Amsterdam: John Benjamins.
- Goddard, C., and Wierzbicka, A. (1994). *Semantic and lexical universals: Theory and empirical findings*. Amsterdam: John Benjamins.
- Goddard, C., and Wierzbicka, A. (2002a). *Meaning and universal grammar – theory and empirical findings*. Amsterdam: John Benjamins.
- Goddard, C., and Wierzbicka, A. Eds. (2002b). *Semantic Primes and Universal Grammar*.

- In C. Goddard & A. Wierzbicka (Eds.), *Meaning and universal grammar – theory and empirical findings*, (pp. 41–85). Amsterdam: John Benjamins.
- Goddard, C., and Wierzbicka, A. (2014). *Words and meanings: Lexical semantics across domains, languages and cultures*. Oxford: Oxford University Press.
- Habib, S. & Kurzon, D. (2008). [The Systematization and Typology of the Instant Messaging Writing System Used by Young Israeli Arabs](#). *Written Language and Literacy*, 11(1), 35-48.
- Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Washington, DC: Georgetown University Press.
- Junger, J. (1987). *Predicate formation in the verbal system of modern Hebrew*. Dordrecht: Foris Publications.
- Kutscher, R. (1982). *A history of the Hebrew language*. Jerusalem: The Magnes Press.
- Levisen, C. (2013). *Cultural semantics and social cognition: A case study on the Danish universe of meaning*. Berlin: De Gruyter Mouton.
- Merriam-Webster, Inc. (2003). *Merriam-Webster's collegiate dictionary*. Springfield, MA: Author.
- Palmer, J. (2007). Arabic diglossia: Teaching only the standard variety is a disservice to students. *Arizona Working Papers in SLA & Teaching*, 14, 111–122.
- Peeters, B. (2006). *Semantic primes and universal grammar: Empirical evidence from Romance languages*. Amsterdam: John Benjamins.
- Peeters, B., Junker, M. O., Travis, C., Farrell, P., Perini-Santos, P., and Maher, B. (2006). Natural Semantic Metalanguage Exponents and Universal Grammar in Romance: Substantives; Determiners; Quantifiers. In B. Peeters (Ed.), *Semantic primes and universal grammar: Empirical evidence from Romance languages*, (pp. 41–78). Amsterdam: John Benjamins.
- Rozen, H. (1962). *A textbook of Israeli Hebrew*. Chicago, London: The University of Chicago Press.
- Ryding, K. C. (2005). *A reference grammar of modern standard Arabic*. Cambridge: Cambridge University Press.
- Saenz-Badillos, A., and Elwolde, J. (1993). *A history of the Hebrew language*. John

- Elwolde (Trans.)*. Cambridge: Cambridge University Press.
- Schiffman, H. (1997). Diaglossia as a Sociolinguistic Situation. In F. Coulmas (Ed.), *The handbook of sociolinguistics*, (pp. 205–16). Oxford, UK; Cambridge, MA: Blackwell Publishers.
- Summers, D., Ed. (2003). *Longman dictionary of contemporary English*. New York: Longman. <http://www.ldoceonline.com/>.
- Walter, E., Ed. (2005). *Cambridge advanced learner's dictionary*. Cambridge: Cambridge University Press.
- Watson, J. (2002). *The phonology and morphology of Arabic*. Oxford: Oxford University Press.
- Wierzbicka, A. (1996). *Semantics: Primes and universals*. Oxford: Oxford University Press.
- Wierzbicka, A. (2002). Semantic Primes and Universal Grammar in Polish. In C. Goddard & A. Wierzbicka (Eds.), *meaning and universal grammar: Theory and empirical findings*, Vol. 2. Amsterdam: John Benjamins.
- Wierzbicka, A. (2009). Exploring English Phraseology with Two Tools: NSM Semantic Methodology and Google. *Journal of English Linguistics*, 37(2), 101–29.
- Wierzbicka, A. (2010). *Experience, evidence, and sense: The hidden cultural legacy of English*. Oxford: Oxford University Press.
- Zuckermann, G. (2004). The Genesis of the Israeli Language: A Response to ‘Philologos’'s ‘Hebrew vs. Israeli.’ *The Mendele Review: Yiddish Literature and Language*, 08.013(152), 1–12.
- Zuckermann, G. (2005). The Israeli Language. *The Mendele Review: Yiddish Literature and Language*, 09.013(165), 1–9.
- Zuckermann, G. (2009). Hybridity versus Revivability: Multiple Causation, Forms and Patterns. *Journal of Language Contact VARIA*, 2, 40–67.