

Chapter 2

Verbal multiword expressions: Idiomaticity and flexibility

Livnat Herzig Sheinfx

University of Haifa

Tali Arad Greshler

University of Haifa

Nurit Melnik

The Open University of Israel

Shuly Wintner

University of Haifa

Verbal multiword expressions are generally characterized by their formal rigidity, yet they exhibit remarkable diversity in their flexibility. Our primary research question is whether the behavior of idioms is an idiosyncratic property of each idiom or a consequence of more general constraints. We challenge Nunberg et al.'s (1994) proposal, attributing decomposability as the determining factor regarding idioms' flexibility/rigidity, first due to the fuzziness of the notion of decomposability, and second, in light of empirical investigations in English and in other languages that revealed flexibility within idioms previously classified as non-decomposable. We propose an alternative classification that builds on the notions of **TRANSPARENCY** and **FIGURATION**. We hypothesize that the more transparent and figurative an idiom is, the more likely it is to be “transformationally productive”. We put this hypothesis to the test by conducting an empirical corpus-based study of a set of idioms of varying degrees of transparency and figuration, using a large corpus of Modern Hebrew.



1 Introduction

Multiword expressions (MWEs) are lexical items that consist of multiple words. They form a heterogeneous class of constructions which include, among others, compounds (e.g., *hot dog*), verb-particle constructions (e.g., *take off*), complex prepositions (e.g., *on top of*), adverbials (e.g., *by and large*) and verbal phrases (e.g., *spill the beans*). MWEs are characterized by their idiosyncratic behavior. The most prominent type of idiosyncrasy ascribed to MWEs is their semantic idiomaticity; their meaning cannot be systematically derived from the meanings that their parts have when they are used independently. For example, there is nothing about the meaning of the words *spill* and *beans* which is necessarily related to the meaning of the idiom *spill the beans*. MWEs may also display idiosyncrasy in other linguistic domains. At the lexical level, MWEs may contain components which are not part of the conventional lexicon (*ad hoc*). Morphologically, they may undergo idiosyncratic processes (*still lifes* and not *still lives* when referring to paintings). Some MWEs have an internal structure which is not accounted for by standard syntax (*by and large*).

MWEs are extremely prevalent: the number of MWEs in a speaker's lexicon is estimated to be of the same order of magnitude as the number of single words (Jackendoff 1997). This may even be an underestimate, as 41% of the entries in WordNet (Fellbaum 1998), for example, are multiwords (Sag et al. 2002). Erman & Warren (2000) found that over 55% of the tokens in the texts they studied were instances of PREFABS (defined informally as word sequences preferred by native speakers due to conventionalization). However, while MWEs constitute significant portions of natural language texts, most of them belong to the long tail in terms of frequency: specific MWEs tend to occur only rarely in texts.

In this chapter we focus on *verbal* MWEs, often referred to as “verbal idioms” or simply “idioms”. Unlike syntactically idiosyncratic expressions such as *by and large*, the structure of verbal idioms is more often than not governed by productive syntactic rules: they contain a verbal head which combines with one or more complements (and possibly adjuncts) to form a verb phrase.¹ Nevertheless, verbal idioms impose stringent selectional restrictions on their lexical components. Moreover, they are known to exhibit “transformational deficiencies” (Chafe 1968: 111), such as resistance to passivization, modification and topicalization. Not all idioms, however, are equally rigid, as some maintain their idiomatic meaning even when they do not appear in their canonical form. The versatile behavior of verbal MWEs raises a question regarding the speakers' knowledge of idioms.

¹The internal structure of some idioms can be syntactically idiosyncratic (e.g., *find fault, close up shop*).

Is information regarding their flexibility encoded for each idiom individually, or can the behavior of idioms be predicted from general principles?

2 Decomposability and flexibility

The groundbreaking work of Nunberg et al. (1994) opened up the possibility of considering the behavior of idioms not as idiosyncratically specified for each idiom individually, but rather as determined by the semantics of the idioms. In this section we first present Nunberg et al.'s (1994) proposal regarding the correlation between the semantic decomposability of idioms and their flexibility/rigidity. We then consider the notion of decomposability and its coherence, and present a number of studies which assessed whether this correlation holds in English and in other languages.

2.1 Decomposability and flexibility: A correlation

The contribution of Nunberg et al. (1994: 503) is set against the background of what they refer to as “well-established assumptions in generative grammar” which is that idioms are non-compositional. In contrast, the authors argue that most idioms do have identifiable parts with assigned interpretations. They distinguish between two types of idioms: *DECOMPOSABLE IDIOMS* (“idiomatically combining expressions” in their terminology) and *NON-DECOMPOSABLE IDIOMS* (“idiomatic phrases”). The former are idioms whose meaning, once known, can be distributed among their parts. A typical example is *spill the beans*, where *spill* roughly means ‘reveal’ and *beans* roughly means ‘secrets’. The meaning of non-decomposable idioms is associated with the *entire* expression; no meanings are assigned to individual words. The often-cited example of this type is *kick the bucket*, for which the meaning ‘to die’ is carried by the phrase in its entirety.

Nunberg et al. (1994) take their analysis a step further by suggesting that there is a correlation between the semantic type of idioms and their behavior. They propose that the semantic distinction between decomposable and non-decomposable idioms accounts for the difference between “transformationally productive” and “transformationally deficient” idioms. The fact that parts of decomposable idioms are assigned interpretations allows them to undergo different “transformations” similarly to ordinary verb phrases.² These parts can be passivized, modified by adjectives or relative clauses, quantified, elided, topicalized/focalized and be antecedents to anaphora. Non-decomposable idioms, on the other hand, only allow verbal inflection.

²We adopt the cover term “transformation” for ease of exposition, with no commitment to its theoretical implications.

2.2 Identifying decomposability

Nunberg et al. (1994) do not provide precise definitions for the two categories, or a specific procedure for distinguishing between them. They do however explicitly warn against confusing decomposability with transparency, which they define as the relation between the literal and idiomatic meaning. Thus, although the idiom *saw logs* is transparent – there is an obvious relation between the sound made by sawing logs and the sound of snoring – it is non-decomposable, since there is no meaning that can be assigned to *logs* in this context. An additional distinction is made between decomposability and paraphrasability. The fact that the meaning of an idiom can be paraphrased using a phrase of a similar argument structure does not necessarily indicate that it is decomposable. For example, although the transitive idiomatic phrase *kick the bucket* could be paraphrased as the transitive phrase *lose one's life* there is nothing about the role of *bucket* in the idiom which suggests that it denotes 'life'.

The coherence of this classification has been put to the test in a number of psycholinguistic experiments. In one experiment Gibbs et al. (1989) compiled a set of idioms which they categorized, based on their own intuitions, into three groups: NORMALLY DECOMPOSABLE IDIOMS for which a part of the idiom is used literally (e.g., *pop the question*), ABNORMALLY DECOMPOSABLE IDIOMS (e.g., *carry a torch*, which refers to the metaphorical extension of torches as warm feelings), and SEMANTICALLY NON-DECOMPOSABLE IDIOMS (e.g., *shoot the breeze*). They presented these idioms along with a paraphrase of their figurative meaning to subjects and asked them to decide whether the individual words in an expression made some unique contribution to its idiomatic meaning, thus testing their intuitions regarding decomposability. As a second step the subjects were instructed to distinguish among the decomposable idioms between those which “have words which are closely related to their individual figurative meaning” (i.e., normally decomposable idioms such as *pop the question*) and those “whose individual words have a more metaphorical relation to their figurative meanings” (i.e., abnormally decomposable idioms such as *spill the beans*).

Gibbs et al. (1989) found that with the exception of three idioms, there was at least 75% agreement among subjects regarding the classification of 36 idioms into one of the three categories. The mean proportion of subject agreement was 86% for those idioms which were initially labeled by the researchers as normally decomposable idioms, 79% for those identified as abnormally decomposable idioms and 88% for semantically non-decomposable idioms. In contrast, Titone & Connine (1994) did not find reliable agreement regarding decomposability in their study. Of the 171 idioms which they examined, only 40% were classified into one

of the three categories (normally decomposable, abnormally decomposable non-decomposable) with at least 75% agreement among subjects. The authors suggest that grouping idioms into these categories may rely on a type of linguistic knowledge that is not easily accessed.

2.3 Empirical assessments of the correlation

Under the assumption that the decomposable/non-decomposable classification is indeed valid, various studies have attempted to assess whether the correlation between decomposability and flexibility holds. Gibbs et al. (1989) presented subjects with idioms in which a lexical item was replaced with a semantically related alternate and with paraphrases of the interpretation of the original idioms. The subjects were asked to judge the similarity between the distorted idiom and the original interpretation. Gibbs et al. found that decomposable idioms were judged by native speakers to be less disrupted by lexical changes. For example, *burst the ice* was found to be more related in meaning to the interpretation of *break the ice* than *kick the pail* was to the interpretation of *kick the bucket*. Similar results were obtained in a set of experiments which focused on syntactic variations (Gibbs & Nayak 1989). Non-decomposable idioms were found to be more limited in terms of the syntactic changes that they can undergo and still retain their figurative meaning. Differences were found also between normally decomposable and abnormally decomposable idioms, where the latter were relatively more constrained in their syntactic behavior. Importantly, not all syntactic operations produced similar results. Some syntactic changes such as adjective insertion and passivization were successful only with normally decomposable idioms. Other changes, such as present participle and adverb insertion, which influence the entire idiom phrase, and not only parts of it, were successful with all types.

A different research method was adopted by Riehemann (2001), who conducted an extensive study of verbal idioms using a 350 million token corpus. She examined four sets of data: (i) idioms that have been discussed in the literature, (ii) idioms that have interesting properties (e.g., passive, negation, adjuncts, no verbal head, more than one idiomatic noun), (iii) idioms with “non-independent words” (or “cranberry expressions”, see Section 3.1 below), and (iv) a random sample of frequent V+NP idioms. Riehemann classified the idioms as decomposable or non-decomposable by attempting to match them with a similarly structured paraphrase. She classified those for which she found an appropriate paraphrase as decomposable. Nevertheless, she observed that the boundary between the two categories is fuzzy. This notwithstanding, her findings show a clear distinction between the decomposable and non-decomposable idioms with respect

to their variability. On average, the canonical forms of idioms account for about 75% of the occurrences of decomposable idioms and 97% of the occurrences of non-decomposable idioms. Moreover, she found that decomposable idioms constitute a majority, with only 27% of the random sample of V+NP idioms classified as non-decomposable.

Nunberg et al.'s (1994) proposed correlation between decomposability and flexibility predicts that non-decomposable idioms would exhibit complete “transformational deficiency”. Nevertheless, Webelhuth & Ackerman (1999) identified a number of German idioms which appear to be non-decomposable yet maintain their idiomatic meaning under topicalization. Bargmann & Sailer (2015) noted similar observations with passivization. Schenk (1995) showed that non-decomposable idioms in German can participate in a verb-second configuration. Verb-second with non-decomposable idioms was also found in Dutch by Grégoire (2007).

Abeillé (1995) argued against the clear bifurcation between fixed and flexible idioms and questioned the validity of the concept of the distribution of the meaning of an idiom among its parts. She examined a sample of 2,000 French verbal idioms and found that most of them did not behave as predicted by Nunberg et al.'s (1994) theory. For example, she showed that a non-decomposable idiom can undergo clefting, provided that a contrastive interpretation, which is a general licensing condition of clefting, is possible.

2.4 Summary

Instances of flexible non-decomposable idioms challenge the all-or-nothing view of transformational deficiencies proposed by Nunberg et al. (1994). Moreover, findings regarding the behavior of idioms in German, Dutch and French cast doubts on the cross-linguistic validity of Nunberg et al.'s (1994) proposal, which was mostly concerned with English idioms. We follow Bargmann & Sailer (2015) in hypothesizing that further research of the flexibility of idioms, especially in languages that differ from English, would reveal language-specific variations that are dependent on language-specific constraints on different transformations.

3 Deconstructing idiomaticity and flexibility

In this chapter we challenge the validity of the hypothesized correlation between decomposability and flexibility. As previously mentioned, decomposability is a fuzzy notion which is difficult to apply when classifying idioms. Although it

was proposed by Nunberg et al. (1994) to be the semantic property of idioms which predicts their behavior, at times this hypothesis is turned around and idiom flexibility is used as a defining property of non-decomposable idioms. Moreover, empirical investigations of the above-mentioned correlation in languages other than English have revealed flexibility within idioms that were classified as non-decomposable. Thus, we argue that decomposability is not a primitive semantic property of idioms, nor can it be used to predict idioms' behavior.

As a first step we picked the quintessential non-decomposable idiom *kick the bucket* to serve as a test case. While this idiom is one of the most frequently cited idioms in the literature, it is scarcely attested in corpora. Moon (1998) did not find any instances of this idiom in the 18 million word corpus that she consulted. Riehemann (2001), using a 350 million word corpus, retrieved only twelve instances, of which one did not appear in the canonical form.

In order to verify that the idiom's common characterization as a rigid idiom is not an epiphenomenon of its low frequency, we consulted a much larger corpus: *enTenTen13* (Baroni et al. 2009), a 20 billion word English corpus, available on SketchEngine (Kilgarriff et al. 2004). Following are examples of determiner variation (1a–1b), modification (1c) and passivization (1d–1e).

- (1) a. When I kick my bucket, Cecelia's yarn can find a new good home.
- b. So what if consuming the foods therein might make us kick that bucket a tad earlier?
- c. My faithful old Samsung i730 PDA phone was starting to kick the battery bucket.
- d. Constantine is a weary, dapper, neo-noir demon-hunting chainsmoker who carries the unfortunate burden of knowing that, when his bucket's kicked, he's going down, not up.
- e. Then Melanie says her last words to Scarlett and falls back onto the starched pillows, her bucket finally kicked.

This preliminary mini-study has shown that given a large enough corpus, even *kick the bucket* can be found to exhibit variations. Consequently, we suggest that the answer to what determines the flexibility or rigidity of idioms is not whether they are decomposable or not. Moreover, we hypothesize that idioms cannot be categorically classified as either flexible or rigid. Rather, we envision a continuum with idioms exhibiting varying degrees of flexibility, possibly dependent on their semantic properties. In an effort to uncover the logic behind the behavior of idioms we reconsider the notions of idiomaticity and flexibility, and propose

an alternative classification, which we then empirically examine by consulting a large corpus of Modern Hebrew.

3.1 Dimensions of idiomaticity

Idiomaticity is often characterized by CONVENTIONALITY; the meaning of idioms cannot be entirely predicted from the meaning of their parts when they appear in isolation from one another. There are, however, a number of other semantic dimensions according to which idioms can be characterized. The dimension which Nunberg et al. (1994: 498) assume plays a crucial role in determining the behavior of an idiom is its decomposability. Nevertheless, as was previously mentioned, determining whether an idiom is decomposable or not is rather impressionistic, and is prone to circularity, where its flexibility is taken as evidence for its decomposability.

In this chapter we consider an alternative categorization of idioms. More precisely, we cross-classify idioms according to two dimensions: FIGURATION and TRANSPARENCY. Figuration reflects the degree to which the idiom can be assigned a literal meaning. Transparency (or opacity) relates to how easy it is to recover the motivation for an idiom's use, or, in other words, to explain the relationship between its literal meaning and its idiomatic one. Idioms are FIGURATIVE if their literal meaning can conjure up a vivid picture in the speaker's mind. Within the figurative idioms we distinguish between two types. In TRANSPARENT FIGURATIVE idioms the relationship between the literal picture and the idiomatic meaning is perceived to be motivated. English examples include *saw logs* ('snore') and *the cat's out of the bag* ('previously hidden facts were revealed'). Conversely, OPAQUE FIGURATIVE idioms portray a picture whose relationship to the idiomatic meaning is not perceptible. English examples include *shoot the breeze* ('chat') and *chew the fat* ('talk socially, gossip'). Idioms which are not figurative do not have a comprehensible literal meaning, and as such are necessarily opaque. Among these idioms we find what are referred to as "cranberry idioms" (Moon 1998; Trawinski et al. 2008), which, similarly to "cranberry morphemes", have parts which have no meanings (e.g., *run amok* 'behave in an unrestrained manner' and *take umbrage* 'take offense'). These idioms may have been figurative and transparent once, but synchronically they contain a word whose meaning is not accessible to contemporary speakers.³

³Opaque non-figurative idioms are not necessarily cranberry idioms. One Hebrew example is *natan ba-kos fein-o* 'gave in the cup his eye' → 'got drunk'. Although all the words in this idiom are common "everyday" words, it does not conjure up any type of image. Such idioms seem to be rare.

In what follows we present a sample of Hebrew idioms representing each of the three categories, namely transparent figurative, opaque figurative and cranberry idioms. Each idiom is illustrated with a corpus example taken from the *heTenTen 2014* corpus (see Section 4.1). We use boldface to highlight the canonical parts of the idioms.⁴ This set of idioms serves as the dataset for our corpus-based investigation of idiom flexibility presented in Section 4.

3.1.1 Transparent figurative idioms

3.1.1.1 *yarad me-ha-ʕec* ‘descended from the tree’ → ‘conceded’

This idiom is part of a more complex expression. To get to a state where a person is required to concede they first need to adopt an unrealistic stance by idiomatically climbing a tall tree: *ʕipes ʕal ʕec gavoha* (‘climbed on tree tall’). Once there, they may need to eventually climb down, or in other words – to concede.

- (2) ʔulai mi ʕe-be-ʕemdat ha-koaħ carix *laredet*
 maybe who that-in-position.cs the-power should *to.descend*
me-ha-ʕec.
from-the-tree

‘Maybe whoever is in a position of power should concede.’

3.1.1.2 *hosif femen la-medura* ‘added oil to the bonfire’ → ‘aggravated the situation’

This metaphorical idiom describes the act of making a situation worse than it already is. A similar English idiom is *add fuel to the fire*.

- (3) beit ha-miʕpaʕ ha-meħozim *hosif femen la-medura:* gam hu
 house.sm.cs the-court the-district *added.3sm oil to.the-bonfire* also he
 lo haya muxan liʕmoʕa ʔet ha-mevaʕef.
 not was ready to.hear ACC the-petitioner

‘The district court added fuel to the fire: it also wasn’t willing to hear out the petitioner.’

⁴The citation form of Hebrew verbs is the third person singular masculine, past tense. Consequently idioms are presented in past tense, and translated as such. When verbs are referred to in isolation their translation is given in the standard English citation form (bare infinitive).

3.1.2 Opaque figurative idioms

3.1.2.1 *ṭaman yad-o ba-calaḥat* ‘buried his hand in the plate’ → ‘refrained from acting’

The origin of this idiom is from the *Book of Proverbs* 19:24, where it describes a person who is so lazy that he leaves his hand in the plate instead of bringing it back into his mouth in order to eat.⁵

- (4) *ṭaman* *ʕacel* *yad-o* *ba-calaḥat* *gam ʔel pi-hu* *lo*
buried.3SM *sluggard.SM* *hand-his* *in.the-plate* *also to* *mouth-his* *not*
yafiv-ena.
 return-it

‘A sluggard buries his hand in the dish; he will not even bring it back to his mouth.’

Most Hebrew speakers are not familiar with the original text and use the idiom in its truncated form. However, without the explicit mention of the actor – the sluggard – and out of context, the idiom is completely opaque, and even more so, it is confusing since it describes an action (i.e., the burying of the hand in the plate), while denoting inaction. Ironically, it is mostly used negatively, to describe someone who does not sit idle.

- (5) *gam be-yaḥasei* *ʔenof lo ṭaman* *yad-o* *ba-calaḥat,*
also in-relations.cs *human not buried.3SM* *hand-his in.the-plate*
pineq *ve-ʔirgen* *lanu micraxim la-piqniq.*
spoiled.3SM *and-organized.3SM* *to.us* *supplies* *to.the-picnic*

‘Also with regards to interpersonal relations, he did not sit idle; he spoiled us and prepared supplies for the picnic.’

3.1.2.2 *heʕela ḥeres (be-yad-o)* ‘brought up a shard (in his hand)’ → ‘tried in vain, failed’

This idiom is figurative since it is possible to imagine someone picking up a shard of clay with their hand. However, it is also opaque since there does not seem to be an obvious relationship between this act and failure. Similarly to the previous idiom, this idiom introduces a paradox: it literally describes the situation of finding something, but it is used to describe an unsuccessful attempt. The original context, unknown to most speakers, is of pearl retrievers who dove in search of

⁵The translation is taken from <http://www.biblestudytools.com/proverbs/19.html>.

pearls, but only came up with a piece of clay (or a pearl-less shell, according to a different interpretation) instead.

- (6) *ha-ʔemet hi ʃe-hipasti gam ʔani ʔax heʃeleti heres*
 the-truth is that-searched.1s also I but brought.up.1s shard
be-yad-i.
in-hand-my

‘The truth is that I also searched, but I failed.’

3.1.2.3 *yaca me-ha-kelim* ‘came out from the tools’ → ‘became upset’

Evidence regarding the opacity of this idiom is found in the ambiguity of the word *kelim*, which could mean ‘tools’, ‘dishes’ or ‘instruments’. There is no consensus among speakers as to which of the meanings applies to this idiom, since none of them seems appropriate. Nevertheless, regardless of the chosen meaning, it is possible to conjure up an image associated with the literal meaning of this expression.

- (7) *hu nirʔa keʔilu hu yoce me-ha-kelim.*
 he looked.3SM as.if he coming.out.SM from-the-tools

‘He looked as if he was becoming upset.’

3.1.3 Opaque non-figurative idioms (cranberry idioms)

3.1.3.1 *ʔavad ʃal-av (ha-)kelaḥ* ‘(the-)KELAH was lost on him’ → ‘became outdated’

The cranberry word in this idiom is *kelaḥ*, which has no known literal meaning. It appears three times in the Old Testament, twice as a name of a place, and once as part of this idiom (Book of Job 30:2). Nevertheless, this idiom is an established part of the Hebrew lexicon. Interestingly, although the noun *kelaḥ* is indefinite in the original Biblical idiom, in Modern Hebrew it is mostly used with a definite prefix.

- (8) *ʔatem ʃosqim be-vikuaḥ ʃe-ʔavad ʃal-av ha-kelaḥ.*
 you engaged-in.argument.SM that-lost.3SM on-him the-KELAH.SM

‘You are engaged in an argument that has become outdated.’

Apart from the cranberry word, one idiosyncrasy exhibited by this idiom is its argument structure. Outside the context of this idiom, the head verb *ʔavad* ‘lose’

does not appear with a PP complement headed by the preposition *ʕal* ‘on’. In addition, unlike the rest of the idioms presented here, this idiom is a full clause, with *kelah* functioning as the subject. Although it is clausal, the complement of the PP is an open slot, and the property of being outdated is predicated on it. Consequently, it is mostly used as a relative clause in which the open slot is occupied by a resumptive pronoun (see also 56).

3.1.3.2 *yafav ʕal ha-meduxa* ‘sat on the MEDUXA’ → ‘deliberated’

The original meaning of the word *meduxa* is ‘mortar’, yet it is not used outside the context of this idiom and its meaning is not known to most Hebrew speakers.

- (9) *ha-cevet yafav ʕal ha-meduxa ve-qiyem kama yefivot.*
 the-team.SM sat.3SM on the-MEDUXA.SF and-held.3SM several meetings
 ‘The team deliberated and held several meetings.’

3.1.3.3 *higdif ʔet ha-seʔa* ‘overfilled the SEAH’ → ‘exaggerated’

The word *seʔa* is originally a biblical unit of measurement, usually of grain, but it is rarely used outside of this idiom (exceptions are texts which deal with religious laws). Interestingly, the verb *higdif* is hardly used outside of this context as well, although the consonantal root *g-d-f* is productive in a different verbal template (*gadaf* ‘fill’).⁶ The original literal meaning of the verb *higdif* was ‘to gather wheat sheaves’, and the literal meaning of the phrase was to overfill a set measure with wheat. The idiom can be used with an agentive subject (10) who “overdoes it”, or an abstract noun (11) which in itself is “too much”.

- (10) *bronil lifʕamim magdif ʔet ha-seʔa be-kama ʕe-hu*
Bronil.SM sometimes overfills.SM ACC the-SEAH.SF in-how.much that-he
meruce me-ʕacmo.
 pleased from-himself
 ‘Bronil sometimes overdoes it in how much he is pleased with himself.’

- (11) *ʕikuv ze higdif ʔet ha-seʔa.*
delay.SM this overfilled.3SM ACC the-SEAH.SF
 ‘This delay was too much.’

⁶Semitic morphology is largely based on roots-and-patterns. Roots are sequences of (typically) three consonants. Patterns are sequences of vowels and possibly consonants with open slots for the roots consonants, indicated by capital Cs. For example, *higdif* and *gadaf* are formed by combining the same consonantal root *g-d-f* with two different templates: hiCCiC and CaCaC.

3.1.3.4 *lo yesula be-paz* ‘will not be SULA in gold’ → ‘priceless’

Unlike the other idioms in this category, the cranberry word in this case is a verb: *sula*. The verb is formed in a passive morphological template (CuCaC) and is never used in the active template (CiCeC). Its original (Biblical) meaning is ‘was measured’, but it is not used out of this context in Hebrew. The noun *paz* ‘gold’ is a very rare synonym of the commonplace *zahav*; its distribution is mostly restricted to fixed phrases (e.g., *hizdamnut paz* ‘golden opportunity’) and the current idiom.

- (12) kenut tihiye kan, ve-ze davar *fe-lo* *yesula*
 honesty.SF will.be.3SF here and-this thing.SM *that-not will.SULA.3SM*
be-paz.
in-gold
 ‘There will be honesty here, and that’s priceless.’

3.2 Types of flexibility

Section 3.1 described semantic dimensions of idiomaticity. In this section we present the formal aspect of this phenomenon, namely the set of lexical, morphological and syntactic transformations that verbal idioms can potentially undergo. We distinguish between four types of transformations: SYNTACTIC VARIATIONS, ARGUMENT STRUCTURE VARIATIONS, LEXICAL INSERTIONS and LEXICAL SUBSTITUTIONS.

3.2.1 Syntactic variations

Syntactic variations are those which preserve the lexical material of the idiom, as well as the grammatical function of the constituents which make up the idiom, but which vary the syntactic configuration of the expression. The occurrence of syntactic variations constitutes evidence against analyses of idioms as fixed phrases (“words with spaces”) which are entered in the lexicon as complete phrases, and which are inserted “as is” into the sentence.

Syntactic variations range from what could be considered as superficial argument shuffling within the VP to extra-phrasal operations such as argument fronting and relativization. One type of syntactic variation that we find is order alternations within the VP. The order of complements in Hebrew is fairly free. Thus, for example, with ditransitive verbs, the position of the two complements can be interchanged with no change of meaning or register.

- (13) a. dan natan matana le-dana.
 Dan gave present to-Dana

- b. dan natan le-dana matana.
 Dan gave to-Dana present
 ‘Dan gave a present to Dana.’

A different case of word order alternation is verb-second. Although the unmarked word order of Hebrew is SVO, subject–verb inversion may be triggered by the occurrence of a clause-initial element, similarly to verb-second constructions. The V2 configuration splits the VP and inserts the subject between the verb and its complements (14b).

- (14) a. * natan dan matana le-dana.
 gave Dan present to-Dana
 b. ?etmol natan dan matana le-dana.
 yesterday gave Dan present to-Dana
 ‘Yesterday Dan gave a present to Dana.’

In addition, we include in the category of syntactic variation two types of long-distance dependencies: topicalization/focalization and relativization. Information structure considerations motivate the fronting of VP-internal material to a clause-initial position. A fronted constituent can be a focal element (15a) or a topicalized element (15b).

- (15) a. [gam le-sara] dan natan matana.
 also to-Sarah Dan gave present
 ‘Dan gave a present also to Sarah.’
 b. [?et ha-matana ha-zot] dan natan le-dana.
 ACC the-present the-this Dan gave to-Dana
 ‘This present, Dan gave to Dana.’

An additional long-distance dependency involves relativization. When NP complements are relativized in Hebrew a resumptive pronoun can optionally occur in the relativization site (16a). Oblique complements are obligatorily resumed by a pronoun, and the language does not allow preposition stranding (16b).

- (16) a. lo ra?iti ?et ha-matana_i [ʃe-dan natan (?ota_i) le-sara].
 not saw.1s ACC the-present that-Dan gave (it) to-Sarah
 ‘I didn’t see the present that Dan gave to Sarah.’
 b. lo ra?iti ?et ha-yalda_i [ʃe-dan natan l-a_i matana].
 not saw.1s ACC the-girl that-Dan gave to-her present
 ‘I didn’t see the girl that Dan gave a present to.’

3.2.2 Argument structure variations

Argument structure variations involve a change in the mapping of arguments to grammatical functions (e.g., passive) as well as valence changing operations (e.g., causativization, reflexivization). Passivization, a primary example of argument structure variation, was found to be a feature that sets decomposable idioms apart from non-decomposable ones (Nunberg et al. 1994), but not in all languages (Bargmann & Sailer 2015).

Hebrew presents an interesting case in this respect, since argument structure variations are associated with the combination of one consonantal root with different morphological templates (Doron 2008; 2003). Thus, the following examples illustrate the root *l-b/v-f* in four different templates: active (17a), reflexive (17b), causative (17c), and passive (17d).

- (17) a. dan lavaf ħulca.
 Dan wore shirt
 ‘Dan wore a shirt.’
- b. dan hitlabef.
 Dan dressed
 ‘Dan got dressed.’
- c. dan hilbij ?et ha-yeled be-ħulca.
 Dan dressed ACC the-child in-shirt
 ‘Dan dressed the child with a shirt.’
- d. ha-yeled hulbaf be-ħulca.
 the-child was.dressed in-shirt
 ‘The child was dressed in a shirt.’

3.2.3 Lexical insertions

Lexical insertions refer to the inclusion of non-selected lexical material within the idiom. This material includes adverbials, quantifiers and different types of noun modifiers. The ability to modify only a part of the meaning is taken by Nunberg et al. (1994) to be a key property of decomposable idioms. They assume that only idiom parts that have individual idiomatic meanings can be modified. However, Ernst (1981) proposed that not all idiom-internal modifiers are equal. He distinguished between INTERNAL and EXTERNAL MODIFICATION, which differ in the semantic scope of their modification. Internal modifiers modify only the element to which they are adjoined (18a). External modifiers, which attach to the object but are semantically associated with the entire verb phrase, are not

indicators of decomposability. The semantically external modifier in (18b) is a “domain delimiter” (Ernst 1981), which specifies the domain to which the idiom applies.

- (18) a. Dan *pulled* some **important** strings. (= Dan used some important connections.)
b. Dan *pulled* some **economic** strings. (= In the domain of economics, Dan pulled some strings.)

The question of whether a modifier is internal or external relates to the issue of decomposability, but not to the question of flexibility. When idiom parts are syntactically modified in an idiom this is certainly an instance of variation; the idiom does not appear in its canonical form. Thus, modifications of all types are variations. The semantic scope of the modifier can provide evidence with respect to whether idiom parts are assigned individual idiomatic meanings or not. For a modifier to be internal the modified part must have its own meaning. For example, *spill royal beans* is acceptable and comprehensible due to the fact that *beans* means ‘secrets’ in the context of this idiom, and royal beans refer to ‘secrets of the royal family’. Conversely, the modifier in *kick the battery bucket* does not attribute any property to the bucket, but rather provides information regarding the domain or cause of death. Nevertheless, the two cases attest to the flexibility of their respective idioms.

3.2.4 Lexical substitutions

Although idioms are known to impose rigid selectional restrictions, there are idioms that maintain their idiomatic meaning even when some of their lexical components are substituted with others. Moon (1998) found verb variation to be the most common type (e.g., *bend/stretch the rules*). Interestingly, the interchangeable verbs are not necessarily synonymous but in that particular context their co-substitution does not alter the idiomatic meaning. As Moon (1998: 50) noted, “searching for verbal variation is the hardest part of corpus-based investigations, and ultimately a matter of serendipity”. For this reason, it is impossible to conduct exhaustive searches of this phenomenon. Nevertheless, with manual inspection of the results of very general queries valuable findings can be gleaned.

4 Corpus findings

In the previous section we proposed an alternative semantic classification of idioms, which builds on the notions of TRANSPARENCY and FIGURATION. In addition,

we distinguished between four types of transformations which idioms can potentially undergo. We hypothesize that the more transparent and figurative an idiom is, the more likely it is to be “transformationally productive”.

To put this hypothesis to the test we conducted an empirical corpus-based study of the behavior of the set of idioms presented in Section 3.1. In light of our preliminary examination of the flexibility of the idiom *kick the bucket*, we chose to consult a very large corpus of Modern Hebrew, which increased the likelihood of finding variations even for relatively infrequently used idioms. The corpus search revealed evidence for variability across the different dimensions of idiomaticity and flexibility. We did not find any idiom which exhibited no variation at all. In what follows we present selected examples of our corpus findings.

4.1 Method

We used *heTenTen 2014* (Baroni et al. 2009), a billion-token web-crawled Hebrew corpus, available on SketchEngine (Kilgarriff et al. 2004), to search for different types of variations that occur with representative MWEs from the three semantic classes we outlined in Section 3.1. We focused on fifteen specific verbal MWEs, and annotated 400 examples overall. Nine of the fifteen MWEs are presented in Section 3.1.

SketchEngine’s Corpus Query Language (CQL) provides a way of defining complex queries which target morphological features of words (e.g., POS, lemmas, clitics) and which make use of logical operators (AND/OR/NOT). These features are particularly important when our goal is to cast a wide net to retrieve variations in general, and in particular non-canonical word orders, discontinuous elements and various morphological inflections. Nevertheless, a wide net comes at a cost; not all the retrieved results are necessarily instances of the idiom. Often, only a manual inspection of each result can weed out the false positives. For this reason we do not present quantitative data with regard to the distribution of the canonical idiom and its variations. We did, however, verify the occurrence of all types of variation for each idiom.

In what follows we present our findings regarding the idioms described in Section 3.1. The presentation is first divided into the four flexibility categories (i.e., syntactic variation, argument structure variation, lexical insertions, and lexical substitutions), and within each category, by the three idiom types (i.e., transparent figurative idioms, opaque figurative idioms, and cranberry idioms). We use boldface to highlight the canonical parts of the idioms, and underline the parts which exhibit the variation under discussion (when possible).⁷

⁷Note that some example sentences exhibit more than one variation, yet in the text we refer only to the one under discussion.

4.2 Syntactic variations

4.2.1 Word order

Verb second configurations are found with all types of idioms. In what follows are examples of transparent figurative idioms (19–20), opaque figurative idioms (21–22), and cranberry idioms (23–24). Note that in all these examples, the clause-initial “trigger” is not part of the idiom. The fronting of idiom parts is discussed in Section 4.2.2.

- (19) *ha-favuša yarad kanir?e misrad ha-datot*
 the-week descended.3SM probably ministry.SM.CS the-religions
me-ha-šec ve-biṭel ?et roša ha-gzera.
from-the-tree and-cancelled.3SM ACC evil.CS the-decree
 ‘This week it seems as if the ministry of religions conceded and cancelled the harsh decree.’
- (20) *yeter šal ken, mosifa ha-fmuša femen la-medura mispar*
 remainder on thus adds.SF the-rumor.SF oil to.the-bonfire few
yamim lifnei qrisat beit ha-hašqašot liman braders.
 days before downfall.CS house.CS the-investment Lehman brothers
 ‘Moreover, the rumor adds fuel to the fire a few days before the downfall of the investment firm Lehman Brothers.’
- (21) *legabei danela hešelu ha-ma?amacim heres.*
 about Danella brought.up.3PM the-efforts.PM shard
 ‘As for Danella, the efforts were unsuccessful.’
- (22) *be-mahalax re?ayon be-yoman ha-caharayim*
 during interview in-daily.news.broadcast.CS the-noon
yaca profesor yosef ?agasi me-ha-kelim ve-amar...
came.out.3SM professor Yosef Agasi from-the-tools and-said.3SM
 ‘During the daily news broadcast at noon, Prof. Yosef Agasi became upset and said ...’
- (23) *be-yamim ?elu yofevet vašadat german šal ha-meduxa.*
 in-days these sitting.SF committee.SF.CS German on the-MEDUXA.SF
 ‘These days, the German committee is deliberating.’

- (24) be-favuša fe-šavar *higdif* fridman *šet ha-seša*.
 in-week that-passed *overfilled*.3SM Friedman.SM ACC *the-SEAH.SF*
 ‘Friedman overdid it last week.’

A different type of word order variation involves an alternative ordering of VP-complements. This, of course, can only occur with ditransitive verbs. Instances of complement ordering with the transitive idiom *hosif femen la-medura* ‘added oil to the bonfire’ → ‘aggravated the situation’ with its canonical lexical parts were not found. Examples (25–26) illustrate this variation, as attested with the opaque figurative idioms *hešela heres (be-yad-o)* ‘brought up a shard in his hand’ → ‘tried in vain, failed’ and *šaman yad-o ba-calaḥat* ‘buried his hand in the plate’ → ‘refrained from acting’, respectively.⁸

- (25) ḥašavti *še-šulai* be-šarxion qaqaš *šemca* tšuvot, *šax gam*
 thought.1s that-maybe in-archive.cs JNF will.find.1s answers but also
šam hešeleti be-yad-i heres.
 there *brought.up.1s in-hand-my shard*
 ‘I thought that maybe I’d find answers in the JNF archive, but I was unsuccessful there as well.’

- (26) šanoxi baḥanti dark-a, ki lo šov, *ve-lo šamanti ba-calaḥat*
 I inspected.1s way-her that not good *and-not buried.1s in.the-plate*
yad-i.
hand-my
 ‘I inspected her behavior, realized it wasn’t good, and did not sit idle (regarding this matter).’

One can note that the canonical form of the cranberry idiom *šavad šal-av (ha-)kelaḥ* ‘(the-)KELAH was lost on him’ → ‘became outdated’ is in a marked VOS order, which is used when the O argument is more topical than the S argument (Melnik 2016). In the following example the idiomatic clause appears in an SVO word order. The placement of a cranberry word in this clause-initial position suggests that it is more topical, which is surprising due to its lack of meaning.

- (27) *ha-kelaḥ* *šoved* gam šal ha-tfisa *še-lefi-ha* kol
the-KELAH.SM is.lost.SM also on the-view.SF that-according.to-her every
ma še-lo racyonali hu šipši.
 thing that-not rational is silly
 ‘The view that anything that isn’t rational is silly is becoming outdated.’

⁸The style of (26) belongs to a high/literary register.

In the following example the PP, which canonically appears between the verb and the subject, appears postverbally, most likely due to “heavy PP shift”.

- (28) *ha-?iʃa hivhira ʃe-ʔavad kelaħ* [ʃal ha-tguvot
 the-woman clarified *that-lost.SM KELAħ.SM on* the-responses
ʃe-ʃolelot ʔet ha-nituħim ha-ʔestetijim mi-kol va-xol].
 that-denounce ACC the-operations the-aesthetic from-all and-all
 ‘The woman clarified that the responses that denounce plastic surgery
 altogether have become outdated.’

4.2.2 Topicalization/focalization of idiom-internal material

Instances of fronting of idiom parts were found across the three semantics types. Following are examples of transparent figurative idioms (29–30), opaque figurative idioms (31), and cranberry idioms (32–34). Note that the fronted part is within square brackets.

- (29) *ʔaval levasof [gam me-ha-ʃec ha-ze] hu yarad.*
 but eventually also *from-the-tree* the-this he *descended.3SM*
 ‘However, eventually he abandoned this stance as well.’
- (30) [*ʔet ha-femen le-medurat ha-geruʃim*] *mosif ʃorex ha-din*
 ACC *the-oil to-bonfire.cs* the-divorce *adds.SM editor.SM.cs* the-law
ʃe-lo ʔovat ha-laqoħot ke-neged ʃein-av ʔela ʔovat heʃbon
 that-not good.cs the-clients as-against eyes-his but good.cs account.cs
ha-banq ʃelo.
 the-bank his
 ‘The lawyer, who is interested in his bank account more than in his
 clients’ interest, adds fuel to the divorce fire.’
- (31) [*gam yedei-hem*] *lo ʔamnu ba-calaxat.*
 also *hands-their not buried.3P in.the-plate*
 ‘Also, they did not sit idle.’
- (32) [*ʃal meduxa ħaʃuva zo*] *yofvim horim rabim.*
 on *MEDUXA.SF* important.SF this.SF *sitting.PM* parents many
 ‘Many parents are deliberating on this important issue.’

- (33) [*ʔet ha-seʔa*] *higdif* ʃadar radio populari...
 ACC *the-SEAH.SF overfilled.3SM* announcer.SM radio popular.SM
 ‘A popular radio announcer overdid it..’
- (34) ʃerk-o ʃacum u-muflag [*ve-be-paz*] lo *yesula*.
 value.SM-its great.SM and-immense.SM *and-in-gold not will.SULA.3SM*
 ‘Its value is great, immense and priceless.’

As is evident from these examples, fronted idiom parts do not usually appear in their canonical form. The fronted constituents in (29) and (31) are modified with the focal marker *gam* ‘also’. In (29) the demonstrative *ha-ze* ‘the-this’ further emphasizes the contrastive interpretation. The topicalized idiom parts in (30) and (32) include modifiers. In fact, in (30) both idiomatic complements are fronted. In examples (32) and (33) the topicalized elements are cranberry words. In (32) the modifier *ħafuva* ‘important’ reveals that *meduxa* is interpreted in this context as the issue on which the parents are deliberating (see more about this in Sections 4.4 and 4.5).

4.2.3 Relativization

When idiom parts are relativized, they surface as the head of the relative clause, which is modified by it. Instances of relativization were found for transparent figurative idioms and cranberry idioms. Examples are given in (35–36) and (37–38), respectively, with the relative clause in square brackets. No instances of relativization were found for opaque figurative idioms. Note that in (36), (37) and (38) a resumptive pronoun inside the relative clause is anaphoric, with an idiom part which functions as the head of the relative clause as its antecedent.

- (35) *ve-nesayem* *be-mila* qṭana ʃal *ha-femen* [*ʃe-hosifu*
and-will.conclude.1P with-word small on *the-oil* *that-added.3PM*
ha-ciyonim la-medura].
the-zionists to.the-bonfire
 ‘And we’ll conclude by mentioning the fuel that the Zionists added to the fire.’
- (36) *ze lo ʃec* *gavoha* [*ʃe-nitan* *laredet* *mime-no*].
 this not *tree.SM* tall.SM *that-possible* *to.descend from-him*
 ‘This is not an unrealistic stance that it is possible to withdraw from.’

- (37) po ʕadayin lo heqimu ʔet *ha-meduxa* [ʃe-ʔeffar *lafevet*
 here yet not established.3P ACC *the-MEDUXA.SF* that-possible *to.sit*
ʕalei-ha ve-ladun ʔeix niftarim me-ha-ʕevra le-meʕeq
on-her and-discuss how get.rid.PM from-the-society to-economy
 ve-kalkala].
 and-economy

‘Here they haven’t yet figured out how to deliberate on the issue of how to get rid of Local Government Economic Services.’

- (38) hu miher la-liʕka, mosad.SM geriyatri.SM *ʃe-ha-kelaħ*
 he hurried to.the-bureau institution geriatric *that-the-KELAH.SM*
 [ʃe-ʔavad ʕal-av] kvar hiħlid mizman.
that-lost.3SM on-him already rusted.3SM long.ago

‘He hurried to the bureau, a geriatric institution that had become outdated long ago.’

The cranberry words in these examples are clearly functioning outside their idiomatic context. In (37) the cranberry word *meduxa* serves as the complement of the verb *heqimu* ‘established’ in a unique and innovative yet comprehensible combination. In (38) *kelaħ* is the head of a relative clause, yet it also functions as a subject of predicate: *kvar hiħlid mizman* ‘rusted long ago’. The speaker in this case attributes to this cranberry word physical properties which are related to aging. In doing so, s/he emphasizes his/her assessment of the bureau as an old and outdated institution.

4.3 Argument structure variations

4.3.1 Mapping variations

Variations with respect to the mapping of idiom parts to grammatical functions were found with transparent figurative idioms. Consider examples (39) and (40). In these examples *femen* ‘oil’, which is the complement of the verb in the canonical idiom, functions as the subject.

- (39) be-xol ʔofen *nifpax* *ha-femen* ʕaxʕav kmo ʃe-ʔomrim
 in-all way was.spilled.3SM *the-oil.SM* now as that-they.say
la-medura.
to.the-bonfire

‘Anyway, fuel was now added, as they say, to the fire.’

- (40) nitan lehosif le-kax ?et ha-gzerot fel ?adri?anus ke-*femen*
 possible to.add to-that ACC the-decrees of Adrianus as-oil.SM
fe-hidliq ?et *ha-medura* fel ha-mered.
that-ignited.3SM ACC *the-bonfire* of the-rebellion
 ‘It is possible to add to that the decrees of Adrianus as fuel which ignited
 the fire of the rebellion.’

In (39) the passive verb *nifpax* ‘was.spilled’ is used instead of the canonical transitive verb *hosif* ‘added’. The agent is not expressed. In (40) the oil is the causer which lights the idiomatic bonfire. Here, too, a different verb is used. This idiom was found to be particularly prone to lexical substitutions (see Section 4.5).

None of the opaque figurative idioms were found to be passivized. Nevertheless, (41) illustrates a different type of argument realization pattern. Indeed, the hand, which is realized as oblique in the canonical idiom (*hefela heres be-yad-o* ‘brought up a shard in his hand’ → ‘tried in vain, failed’) is realized in this case as the subject.

- (41) zar ha-mefafpef be-ma?agarei hasfarim
 outsider.SM that-rummaging.SM in-stock.CS the-books
 yufta legalot ki *yad-o* *mašala* *heres*.
 will.be.surprised.3SM to.discover that *hand.SF-his* brings.up.SF shard
 ‘An outsider rummaging in the stock of books would be surprised to find
 himself unsuccessful.’

In the cranberry idiom *higdif ?et ha-se?a* ‘overfilled the SEAH’ → ‘exaggerated’ the cranberry word *ha-se?a* functions as the complement of the verb. There are, however, instances of this idiom where *ha-se?a* functions as the subject and the verb is a morphological variant of the canonical verb: the passive *hugdefa* ‘was.overfilled’ in (42) and the middle *nitgadfa* ‘was.overfilled’ in (43).

- (42) ha-pašam **hugdefa** *ha-se?a*.
 the-time was.overfilled.3SF *the-SEAH.SF*
 ‘This time things went overboard.’
- (43) Šata **nitgadfa** *ha-se?a* ve-higiša ha-šet letaqen ?et
 now was.overfilled.3SF *the-SEAH.SF* and-arrived the-time to.fix ACC
 ha-mešuvat.
 the-wrong
 ‘Now that things went overboard, it is time to fix the wrongdoing.’

4.3.2 Causativization

The verbs *yarad* ‘descend’, *yaca* ‘come out’ and *yafav* ‘sit’, which head a transparent figurative, opaque figurative and cranberry idiom, respectively, are also used in their causative form in the same idioms. The causee argument, indicated by square brackets, is an “open slot”. This is illustrated in (44), (45) and (46).

- (44) *ʔein kmo ʔilot biʔhonyot lehorid* [ʔet ʔnei ha-cdadim]
 no like reasons security **to.bring.down** ACC both.cs the-sides
me-ha-ʔec.
from-the-tree

‘There is nothing like security reasons to cause the two (opposing) sides to concede.’

- (45) *ha-ʔofanayim ha-ʔele yexolim lehoci* [ʔafilu baʔur ʔaqeʔ kamoni]
 the-bicycle.PM the-these can.PM **to.take.out** even guy quiet like.me
me-ha-kelim.
from-the-tools

‘This bicycle can make even a quiet guy like me upset.’

- (46) *higiʔa ha-zman lehofiv ʔal ha-meduxa* [kalkelanim
 arrived the-time **to.cause.sit** on *the-MEDUXA.SF* economists
ve-aqʔuʔarim].
 and-actuaries

‘It’s time to make economists and actuaries deliberate (on some issue).’

The following example exhibits a neologism. The verb *heʔavid* is created by combining the consonantal root of the original verb *ʔavad* ‘lose’ (*ʔ-b/v-d*) with the causative template HiCCiC to create a verb whose meaning is ‘cause to be lost’.⁹ The cranberry word *kelaʔ* serves as the causee and surfaces as a direct object (marked with the accusative case marker *ʔet*). This suggests that it is interpreted (at least in this case) as some property, perhaps relevance, whose absence makes something outdated. Although this neologism is attested only once in the corpus, it is comprehensible in the context of this idiom, due to the transparent morphological relationship between it and the canonical verbal form.

⁹The original Biblical meaning of *heʔavid* is ‘demolish, destroy’.

- (47) *sifro ha-rifon doħe ?et ha-hafqafa ha-rovaħat fe-qant*
 his.book the-first rejects ACC the-view the-common that-Kant.SM
he?evıd ?et ha-kelah řal ha-meřafızıqa.
cause.be.lost.3SM ACC the-KELAH.SM on the-metaphysics
 ‘His first book rejects the common view that Kant made metaphysics
 outdated.’

4.4 Lexical insertions

Transparent figurative idioms were found to be amenable to different types of modifications. Consider the following examples with the idiom *yarad me-ha-řec* ‘descended from the tree’ → ‘conceded’.

- (48) *hi crixal laredet me-ha-řec ha-cadqani*
 she needs.SF to.descend from-the-tree.SM the-righteous.SM
ve-ha-baxyani fe-řal-av hi řıpsa.
and-the-whiny.SM that-on-him she climbed.3SF
 ‘She needs to abandon the righteous and whiny stance which she
 adopted.’
- (49) *?oto řef mesaper fe-hu himci ?et ha-řuga ha-popolarit. yef*
 same chef tells that-he invented ACC the-cake the-popular there.are
ka?elu fe-holqım řal-av ve-menasım lehorıd ?oto qcat
 those.PM that-disagree.PM on-him and-try.PM to.bring.down him bit
me-ha-řec ha-řoqoladı.
from-the-tree.SM the-chocolaty.SM
 ‘The same chef claims that he invented the popular (chocolate) cake.
 There are those who disagree and try to make him slightly abandon his
 stance re. chocolate.’

In example (48) the idiomatic tree is modified by two adjectives, and with a relative clause which includes the associated idiom *řıpes řal řec gavoha* ‘climbed on a tall tree’ → ‘adopted an unrealistic stance’. In the canonical idiom the adjective *gavoha* ‘tall’ modifies *řec* ‘tree’ by relating to its literal sense, yet in this example the adjectives modify the assumed idiomatic meaning of *řec* ‘tree’ → ‘stance’. The modifier *řoqoladı* ‘chocolaty’ in (49) also modifies *řec* ‘tree’, but in this case it is an external “domain delimiter” (Ernst 1981), which specifies the domain to which the idiom applies.

A different type of variation is exhibited by the following example, where the idiomatic tree is both quantified and pluralized. In addition, the adverb *befeqet* ‘quietly’ externally modifies the entire phrase.

- (50) *kulam yordim befeqet mi-kol ha-secim.*
 everyone *descending.PM* quietly from-all *the-trees*
 ‘Everybody is quietly abandoning all of their stances.’

Opaque figurative idioms exhibit less variation in terms of different types of modification. Following are two examples.

- (51) *gam gugel lo tomenet ?et yad-a ha-virtu?alit ba-calahat.*
 also Google.SF not burying.SF ACC hand.SF-her the-virtual.SF in.the-plate
 ‘Google also isn’t sitting idle.’

- (52) *heSeleti be-yad-i heres muhlat.*
 brought.up.1s in-hand-my shard.SM absolute.SM
 ‘I absolutely failed.’

The adjective *virtu?alit* ‘virtual’ in (51) is a domain delimiter, similarly to *foqoladi* ‘chocolate’ in (49) above. The modification of *heres* ‘shard’ in (52) is also external: the speaker describes a situation where she searches for something and finds *absolutely* nothing.

The occurrence of modification in cranberry idioms is especially surprising due to their opaqueness and lack of figuration. Nevertheless, as the following examples show, cranberry words are compatible with different types of modifications. In (53) the cranberry word *meduxa* is modified by an adjectival phrase and with a demonstrative, in (54) by a relative clause and in (55) *meduxa* is the head of a construct state NP which is modified by its complement. In all three instances, the modification suggests that the speakers perceive the interpretation of *meduxa* to be ‘issue’. A similar case was presented in (32) above.

- (53) *mif?ala codeqet ve-nexona zo ha-meqanenet be-lev kol*
 wish justified and-right this that-lies in-heart.cs every
ha-yofvim fal meduxa qafa ve-tragit zo...
that-sitting.PM on MEDUXA.SF difficult.SF and-tragic.SF this.SF
 ‘This justified and right wish, that lies in the heart of all those deliberating on this difficult and tragic issue..’

- (54) notnim le-fofet *lafevet šal meduxa* **je-lo qayemet**, ve-lehaxlit
 give to-judge *to.sit on MEDUXA.SF* **that-not exists.SF** and-to.decide
 je-yef latet lahem ?et ?admot ha-negev *šelanu*.
 that-must to.give to.them ACC plots.cs the-negev ours
 ‘(They) assign a judge to deliberate on a non-existing issue, and decide
 that our plots in the Negev should be given to them!’

- (55) *šal meduxat heter šimuf be-nešeq le-našim* kvar *yafvu*
on MEDUXA.SF.CS license use.CS in-weapon to-women already *sat.3PM*
 posqei dorenu.
 adjudicators.PM.CS our.generation
 ‘The adjudicators of our time have already deliberated on the issue of
 women using weapons.’

Instances of lexical insertions with the other cranberry idioms were also found. In (56) the cranberry word *kelah* appears with an indefinite quantifier and an adjective, although it is not clear what it denotes, neither literally nor idiomatically. In (57) the cranberry word *se?a* is modified with a relative clause, which refers to its literal meaning as a measure.

- (56) hem hoqrim *je-miftamšim be-ṭexniqot ve-tfisot* šolam
 they researchers that-use in-techniques.PF and-views.PF.CS world
je-?avad šalei-hem ?eize kelah qaṭan.
that-lost.3SM on-them some KELAH.SM small.SM
 ‘They are researchers who use techniques and world-views that are a bit
 outdated.’
- (57) be-cašad ze *higdaštem ?et ha-se?a ha-mele?a gam kax*
 in-step this *overflowed.2PM ACC the-SEAH.SF the-full.SF* also *this.way*
be-pigušei ṭeror.
in-attacks.cs terror
 ‘With this step you overdid a situation that was already too much with
 respect to the terror attacks.’

4.5 Lexical substitutions

4.5.1 Transparent figurative idioms

Note that transparent figurative idioms exhibit lexical substitution of both verbs and nouns. In (58) the verb *yarad* ‘descend’ is replaced with the more active *qafac* ‘jump’, without loss of idiomatic meaning.

- (58) hu hevin fe-cadaqnu ve-maca derex mavriqa
 he understood.3SM that-were.right.1P and-found.3SM way brilliant
liqqoc *me-ha-ʕec.*
to.jump *from-the-tree*

‘He understood that we were right and found a brilliant way to concede.’

The idiom *hosif femen la-medura* ‘added oil to the bonfire’ → ‘aggravated the situation’ exhibits lexical substitutions of both the verb and the noun. The verb *hosif* ‘add’ is substituted by different verbs whose meaning approximates ‘adding something (mostly liquid) to something else’. One such case is exemplified in (59). Other examples are *nifpax* ‘be spilled’ in (39) and *hidliq* ‘ignite’ in (40).

- (59) wikipedia **yocqet** *femen la-medura* ʕel hiluqei
 Wikipedia.SF **pouring.SF** *oil* *to.the-bonfire* of disagreements.cs
 ha-deʕot ha-beinleʔumiyim.
 the-opinions the-international

‘Wikipedia is seriously aggravating the situation of international disagreements.’

More creative variations are found with the substitution of nouns. In (60) *femen* ‘oil’ is substituted by the more general *homer nafic* ‘explosive material’, still within the semantic domain of the literal meaning. In (61) it is replaced with an abstract noun which refers to its idiomatic interpretation.

- (60) ʔanaħnu lo crixim leħapes ʕod **homer nafic** *lehosif*
 we not need to.search more **material explosive** *to.add*
la-medura.
to.the-bonfire

‘We don’t need to search for additional ways to aggravate the situation.’

- (61) mazuz ʕacmo *hosif* *la-medura* ʔet ha-hitnahalut
 Mazuz.SM himself *added.3SM to.the-bonfire* ACC **the-conduct**
 ha-ʕaʕaruriyatit be-tiq qacav.
 the-outrageous in-case.CS Katsav
 ‘Mazuz himself aggravated the situation, with the outrageous conduct in Katsav’s case.’

4.5.2 Opaque figurative idioms

Lexical substitutions are also found in the category of opaque figurative idioms. In (62) the idiom is exploited to describe not the act of bringing up a piece of shard, but rather the end result: remaining with it in your hand. This change of verb and perspective does not disrupt the meaning of the idiom.

- (62) ʔazai kʕe-tagifa kvar le-bhinat ha-hitqadmut
 then when-will.arrive.2SM already to-examination.cs the-step
 ha-hamcaʔatit tivater ʕim *heres be-yad-xa*.
 the-inventive will.remain.2SM with *shard in-hand-your*
 ‘When you finally get to the examination of inventive step you will have failed.’

In (63) and (64) the hand which in the canonical idiom is buried (or not) in the plate, is replaced with other instruments: a camera and a sting. These expressions can only be understood provided that the canonical idiom is known.

- (63) gam nadav *lo ʕaman* **maclema-to** *ba-calaħat*
 also Nadav.SM *not buried.3SM camera-his in.the-plate*
 ve-cilem ba-mozeʔon lelo heref.
 and-photographed.3SM in.the-museum without stop
 ‘Nadav also didn’t refrain from using a camera, and took pictures in the museum without stopping.’
- (64) gam ha-cirʕa *lo ʕomenet ʔet ʕoqc-a* *ba-calaħat*.
 also the-wasp.SF *not bury.SF* ACC **sting-her** *in.the-plate*
 ‘The wasp also does not refrain from using its sting.’

4.5.3 Cranberry idioms

Cranberry idioms are also subject to lexical substitutions. There are a few instances of the idiom *lo yesula be-paz* ‘will not be SULA in gold’ → ‘priceless’

where the cranberry verb *sula* is replaced with a Hebrew synonym *heferix* ‘evaluate’. One such example is given in (65).

- (65) ze mifgaf hevrati fe-leʔolam lo yahzor ve-ʔein
 this get.together.SM social.SM that-never not will.return.3SM and-not
 lehaʔarix-o be-paz.
 to.evaluate-him in-gold
 ‘This is a social get-together that will never return and should be considered priceless.’

In (66) the verb *yafav* ‘sit’ is substituted by the verb *hitqabec* ‘gather’, yet the idiomatic meaning is maintained.

- (66) ʔal meduxa zo hitqabcu harbe melumadim.
 on MEDUXA.SF this.SF gathered.P many scholars
 ‘Many scholars have deliberated on this issue.’

4.6 Discussion

Verbal MWEs in Hebrew turned out to be consistently more flexible than would be expected given Nunberg et al.’s (1994) categorical bifurcation. All the idioms we investigated in this study exhibited flexibility to a certain extent. The variations exhibited by the transparent figurative idioms refer to both the literal and the figurative meanings of the expressions. Thus, speakers can relate to the tree in *yarad me-ha-ʔec* ‘descended from the tree’ → ‘conceded’ in its literal meaning as an entity with physical properties (e.g., ‘tall’ in 36) which can be physically manipulated, either by climbing down from it (in the canonical form) or by jumping down from it (58). The height of the tree or the manner with which one descends from it transfer metaphorically to the idiomatic meaning of the phrase. Conversely, speakers can also attribute to the tree in the idiom abstract properties which are only appropriate in the context of the idiom (e.g., *cadqani ve-baxyani* ‘righteous and whiny’ in 48).

Even more flexibility is found with the transparent figurative idiom *hosif femen la-medura* ‘added oil to the bonfire’ → ‘aggravated the situation’. The vivid picture which this idiom conjures allows speakers to describe it in different terms, while still maintaining the idiomatic meaning. Thus, we find lexical substitutions for both the verb *hosif* ‘add’ and the noun *femen* ‘oil’, which refers to the material added to the bonfire (both literal as in (60) and idiomatic as in (61)). As far as we can tell, the word *medura* ‘bonfire’ cannot be substituted.

On the other side of the flexibility continuum are the opaque figurative idioms. While different types of variations were found to be compatible with these idioms, this class exhibited a more constrained behavior. We did not find any evidence for instances of relativization, and only a handful of cases of lexical insertions. Lexical substitutions, too, were rare. The two examples given for *taman yad-o ba-calahaṭ* ‘buried his hand in the plate’ → ‘refrained from acting’ in (63) and (64) are instances of what could be considered as “word play”. Furthermore, the opacity of this idiom is especially evident in light of attested instances where its use reflects a wrong/alternative interpretation of the idiom, one in which the burying of the hand indicates involvement in something. This is illustrated in (67).

- (67) *ʔarbaʕim ḥavarot beinleʔumiyot ʔomnot ʔet yad-an*
 forty companies.PF international.PF bury.PF ACC hand-their
be-calahaṭ ha-zihum ha-gdufa ʕel sin.
in-plate.SF.CS the-pollution the-full.SF of China
 ‘Forty international companies are involved in heavily polluted China.’

We suggest that the combination of figuration and opacity emphasizes the idiosyncrasy of these idioms, and consequently speakers are more conservative in the way that they use them.

We were especially surprised by the behavior of the cranberry idioms. Our initial expectation was that the lack of transparency and figuration would render these idioms more rigid. Our corpus findings, however, reveal a different picture. The usage patterns exhibited by these idioms suggest that speakers attribute to the meaningless cranberry words some semantic content, or to put it more idiomatically – breathe new life into them. As was illustrated and discussed above, the usage patterns of these idioms suggest that speakers are imposing some interpretation on cranberry words. The word *meduxa* in *yafav ʕal ha-meduxa* ‘sat on the MEDUXA’ → ‘deliberated’ is interpreted as denoting the issue which is under deliberation (see 32, 53 and 55). A similar situation is found with respect to *kelah*. From the examples, we can see that in spite of its lack of meaning it is conceptualized as a physical object which can be small (56) and can become rusty (38). Moreover, it can function as the topic of a clause (27 & 38). It would seem that the meaninglessness of cranberry words frees speakers to apply their own interpretation to them, and to provide idioms which are opaque and non-figurative with transparency and figuration.

4.7 Conclusion

In this chapter we challenged the predictive ability attributed to the notion of decomposability by Nunberg et al. (1994). We argued that this notion is too fuzzy to be used as a principle for reliably categorizing idioms, and, moreover, that it cannot be used to predict their flexibility or rigidity. On the contrary, we hypothesized that idioms cannot be categorically classified as either flexible or rigid, rather, that they occupy a continuum, with different idioms exhibiting varying degrees of flexibility.

We questioned the validity of the assumption that some idioms are completely rigid (modulo verbal inflection) and demonstrated that even the quintessential non-decomposable idiom *kick the bucket* can undergo transformations. However, since this idiom is used relatively rarely and idiom variations in and of themselves are relatively infrequent, non-canonical instances of it and other infrequent idioms can only be empirically attested in very large corpora. This, we believe, is an important methodological finding, which at this point in time, with the availability of large annotated corpora, cannot be overlooked.

Rather than focusing on decomposability as a defining property of idioms, we considered two distinct semantic dimensions: FIGURATION and TRANSPARENCY. We hypothesized that the more figurative and transparent an idiom is, the more amenable it is to various transformations. Our corpus-based investigation and subsequent comparison of idioms associated with three semantic types (transparent figurative, opaque figurative and opaque non-figurative) revealed that the usage patterns of opaque figurative idioms are the most conservative among the three.

Opacity, however, was found not to be the sole “culprit”, since cranberry idioms which contain meaningless words were found to be relatively flexible. Thus, we propose that neither transparency nor figuration alone can account for the behavior of idioms. Our findings suggest that there is an interaction between the two dimensions. Figurative idioms are flexible dependent on their transparency: when transparent they are relatively amenable to various transformations. Conversely, the flexibility of opaque idioms depends on their figuration: when opaque idioms are not figurative due to the inclusion of meaningless cranberry words speakers can ascribe to these meaningless words content which renders the idioms more figurative and less opaque, and consequently – more flexible. Naturally, this generalization, which is based on our work on only a limited set of Hebrew verbal idioms, requires further investigation.

Abbreviations

1/2/3	person	ACC	accusative case
S/P	number	CS	construct state
F/M	gender		

Acknowledgements

This research was supported by The Israel Science Foundation (grant No. 505/11).

References

- Abeillé, Anne. 1995. The flexibility of French idioms: A representation with lexical tree adjoining grammar. In M. Everaert, E. van der Linden, A. Schen & R. Schreuder (eds.), *Idioms: Structural and psychological perspectives*, chap. 1, 15–42. Hillsdale: Lawrence Erlbaum Associates.
- Bargmann, Sascha & Manfred Sailer. 2015. *Syntactic flexibility of non-decomposable idioms*. Poster at the 4th General PARSEME Meeting, Valletta, Malta.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226. DOI:10.1007/s10579-009-9081-4
- Chafe, Wallace L. 1968. Idiomaticity as an anomaly in the Chomskyan paradigm. *Foundations of language* 4(2). 109–127.
- Doron, Edit. 2003. Agency and voice: The semantics of the Semitic templates. *Natural Language Semantics* 11(1). 1–67.
- Doron, Edit. 2008. The contribution of the template to verb meaning. In Galia Hatav (ed.), *Modern linguistics of Hebrew*, 57–88. Jerusalem: Magnes Press. In Hebrew.
- Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20(1). 29–62.
- Ernst, Thomas. 1981. Grist for the linguistic mill: Idioms and ‘extras’ adjectives. *Journal of Linguistic Research* 1(3). 51–68.
- Fellbaum, Christiane (ed.). 1998. *WordNet: An electronic lexical database* (Language, Speech and Communication). MIT Press.
- Gibbs, Raymond W. & Nandini P. Nayak. 1989. Psycholinguistic studies on the syntactic behavior of idioms. *Cognitive psychology* 21(1). 100–138.

- Gibbs, Raymond W., Nandini P. Nayak, John L. Bolton & Melissa E. Keppel. 1989. Speakers' assumptions about the lexical flexibility of idioms. *Memory & Cognition* 17(1). 58–68.
- Grégoire, Nicole. 2007. Design and implementation of a lexicon of Dutch multiword expressions. In *Proceedings of the workshop on a Broader Perspective on Multiword Expressions*, 17–24. Prague, Czech Republic: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W/W07/W07-1103>.
- Jackendoff, Ray. 1997. *The architecture of the language faculty*. Cambridge, MA, USA: MIT Press.
- Kilgarriff, Adam, Pavel Rychly, Pavel Smrz & David Tugwell. 2004. The Sketch engine. In *Proceedings of EURALEX*, 105–116. Lorient, France.
- Melnik, Nurit. 2016. *The development of the VOS construction*. Paper presented at the “Language contact, language continuity and change, and the emergence of Modern Hebrew” workshop, Hebrew University, Jerusalem.
- Moon, Rosamund. 1998. *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Oxford University Press.
- Nunberg, Geoffrey, Ivan Sag & Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538.
- Riehemann, Susanne. 2001. *A constructional approach to idioms and word formation*. Stanford University dissertation.
- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the 3rd international conference on Computational Linguistics and Intelligent Text Processing* (Lecture Notes in Computer Science 2276), 1–15. Springer.
- Schenk, André. 1995. The syntactic behavior of idioms. In M. Everaert, E. van der Linden, A. Schen & R. Schreuder (eds.), *Idioms: Structural and psychological perspectives*, 253–272. Hillsdale: Lawrence Erlbaum Associates.
- Titone, Debra A. & Cynthia M. Connine. 1994. Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literalness. *Metaphor and Symbol* 9(4). 247–270.
- Trawinski, Beata, Manfred Sailer, Jan-Philipp Soehn, Lothar Lemnitzer & Frank Richter. 2008. Cranberry expressions in English and in German. In *Proceedings of the LREC workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*, 35–38.
- Webelhuth, Gert & Farrell Ackerman. 1999. A lexical-functional analysis of predicate topicalization in German. *American Journal of Germanic Linguistics and Literatures* 11(01). 1–61.