



The Prague Bulletin of Mathematical Linguistics
NUMBER 118 APRIL 2022 25-53

Hebrewnetette - A New Derivational Resource for Non-concatenative Morphology: Principles, Design and Implementation

Lior Laks,^a Fiammetta Namer^b

^a Bar-Ilan University - Ramat-Gan, Israel

^b UMR 7118 ATILF, Université de Lorraine & CNRS - Nancy, France

Abstract

This paper presents the architecture of a derivational database of Modern Hebrew (and more generally of Semitic languages) called Hebrewnetette. The methodology adopted is based on adjusting the structure and properties of a database developed for the description of the derivational relations in the lexicon of a Romance language (Démonette), and providing it with additional features to account for the specificities of the morphology of Semitic languages, with special reference to root-and-pattern non-concatenative morphology. We present the properties of Hebrewnetette and the type of information it consists of, with emphasis on both structural and semantic relations between words. We show how this is implemented and examine two case studies, where we demonstrate how the annotations that are used allow us to verify theoretical hypotheses about non-concatenative morphology. The design of Démonette's annotation system allow its features, initially designed for French, to capture morphological and semantic relations between Hebrew words, regardless of the type of morphology (concatenative or non-concatenative).

1. Introduction

This paper presents a derivational resource for Modern Hebrew based on an existing infrastructure that was originally designed for Romance languages like French. We show how the existing architecture of the database can be adapted for Semitic morphology with some relevant additions.

Current available resources and tools for European languages can be divided in two main types:

- The first ones specifically describe a process (or a family of processes) of a given language. The reader can refer to Kyjánek (2018) for a typological description of the structure and coverage of 30 recent derivational resources for Romance (including Latin), Germanic and Slavic languages, which provides a complete list of the main existing derivational databases and lexicons with derivational annotations.
- The second type of databases aims at a multilingual description by homogenizing linguistically and structurally heterogeneous sources: construction of standards in terms of tagsets (McCarthy et al. 2018), standardization of existing resources in the form of architectures with a universal vocation (Universal Derivation (Kyjánek et al. 2020), MorphyNet (Batsuren et al. 2021), UniMorph (Kirov et al. 2018)). This second category of databases is fed (among other things) by the content of the first ones.

This article is devoted to the first of these two categories of databases. Specifically, we ask how a database designed and developed to represent the derivational properties of French - and more broadly, of a Romance language - can be used to describe the morphology of Semitic languages, and more particularly the non-concatenative derivational relations. The database we use as a starting point for this study is *Démonette* (Namer and Hathout 2020). In order to address this issue, we break it down into three questions.

- What is needed, in the design of a database, to represent in a satisfactory, exhaustive and fine-grained way the derivational properties of the non-concatenative (as well as concatenative) morphology of Semitic languages?
- Since *Démonette* is designed for the fundamentally concatenative morphology of a language like French, thus genotypically very distant, could its principles be applied for this purpose, by means of additional descriptions but without modifying the existing structure? If yes, this can serve as an initial prototype for a unified framework for the description of the derivational morphology of many languages.
- Does the database succeed in representing theoretical issues regarding Semitic Morphology?

To answer these three questions, our paper is structured as follows. In Section 2, we recall the morphological principles that distinguish Hebrew (and Semitic languages in general) from Western European languages. Then, we present *Démonette*, the derivational database of French that we adapt to Hebrew (Section 3). Section 4 describes the extensions (new attributes, new values) that formalize the specific derivational properties of Hebrew in the *Hebrewnette* database, while preserving the architecture of the source database and keeping the original features and values. Section 5 presents two case studies, which put the *Hebrewnette* feature structure to the test.

Finally, in Section 6, we explain how we chose the entries in order to build the current version of Hebrewnet, with the aim of evaluating the relevance of the proposed tagset, and assessing its capability of covering a wide variety of types of structural and semantic relations in Hebrew.

2. Hebrew Morphology

2.1. Root-and-pattern Morphology

Hebrew word formation relies highly on non-concatenative morphology, i.e. via root and pattern (Aronoff 1994, Berman 1978, Bolozky 1978, Ravid 1990, Schwarzwald 1981). The pattern indicates the prosodic structure of the word and it consists of the following elements: (i) consonantal slots; (ii) vocalic pattern; and in some cases (iii) affixes (Bat-El 1994, 2017). For example, the verbs *siper* ‘tell_V’ and *limed* ‘teach_V’ are formed in the CiCeC pattern.¹ They share the vocalic pattern i-e and differentiate in their roots, s-p-r and l-m-d respectively. The verbs *hitkavec* ‘become shrunk_V’ and *hitravev* ‘become wide_V’ are formed in the hitCaCeC pattern, which consists of the prefix *hit-*, in addition to the vocalic pattern a-e.

2.1.1. Verb patterns

Words that share the same consonantal root typically share some semantic relations with different degrees of transparency, for example *hidpis* (hiCCiC) ‘print_V’, *hudpas* (huCCaC) ‘be printed_V’, *madpeset* (maCCeCet) ‘printer_N’ and *tadpis* (taCCiC) ‘printout_N’. Hebrew verbal patterns typically differ from each other with respect to transitivity and the semantic types of verbs that they host (see Aronoff 1994, Berman 1978, Bolozky 1978, Borer 1991, Doron 2003, Ravid 1990, Ravid et al. 2016: and references therein). For example, CiCeC typically hosts active transitive verbs, e.g. *kivec* ‘shrink’, *nigev* ‘wipe’ and *xibek* ‘hug’, while hitCaCeC typically hosts intransitive verbs like inchoatives (*hitkavec* ‘become shrunk’), reflexives (*hitnagev* ‘wipe oneself’) and reciprocals (*hitxabek* ‘hug each other’). However, these only represent tendencies and there is no one-to-one correspondence between form and meaning of the patterns. For example, *hitpalel* ‘pray’ is formed in hitCaCeC but does not belong to any of the above mentioned semantic classes.

Within verb formation, non-concatenative formation is obligatory, and every verb that enters the language must conform to one of the existing patterns. This is attested in the formation of verbs that are derived from words without roots, including words borrowed from various languages. For example, the verbs *midel* ‘make a model (out

¹The term “formed” indicates that a specific word shares the form of one of the existing patterns. It does not necessarily imply that a word formation pattern actually took place. Rather, it denotes the fact that (a) the word has the vocalic melody and the affixes (if any) of that pattern, and (b) the root consonants of this word occupy the consonantal slots of the pattern.

of smth)' (CiCeC), *hisvim* 'send a spam' (hiCCiC) and *hitfakes* 'be in focus' (hitCaCeC) are derived from English loan words and are formed in three different patterns. Most transitive verbs are formed in CiCeC by default and some are formed in hiCCiC in order to preserve the consonant cluster of the base (*spam-hisvim*). The verb *hitfakes* is formed in hitCaCeC because it is an intransitive verb. Newly coined intransitive verbs that are inchoative, reflexive and reciprocal are formed in hitCaCeC almost exclusively.

2.1.2. Nominal and adjectival patterns

Hebrew also has a set of patterns that are used for the formation of nouns and adjectives. Most patterns have typical meanings, although meanings that are associated to them represent mainly tendencies. For example, the maCCeC pattern is typical of instrument nouns like *masnen* 'filter' and *masrek* 'comb', but is used for the formation of other nouns, e.g. *martef* 'basement'. The CaCaC pattern is typical of agent nouns (e.g. *cayar* 'painter'), but is also used for instrument nouns (e.g. *vasat* 'regulator') and adjectives (e.g. *raxav* 'wide').

Each verbal pattern, apart from the passive patterns CuCaC and huCCaC, is related to a typical nominal pattern that is used for the formation of actions nouns. For example, the typical action noun pattern of CaCaC is CCiCa (e.g. *kataw* 'write' - *ktiv* 'writing') and that of CiCeC is CiCuC (e.g. *limes* 'teach' - *limud* 'teaching'). There is some interpredictability between the verbal and nominal patterns, and this allows us to feed the database in a semi-automatic way (see Section 6). However, this system is subject to a certain amount of irregularity. For example, the action noun of *higer* (CiCeC) 'emigrate' is *hagira* and not **higur*. Moreover, some of the action nouns have an additional nominal meaning. This is a well-known action-result polysemy phenomenon, where the deverbal action noun also denotes the result of such action. Such polysemy can be found in many languages, as has been shown in various studies (see, among others, Alexiadou 2001, Berman and Seroussi 2011, Borer 2014, Comrie and Thompson 2007, Grimshaw 1990, Hazout 1995, Levin and Rappaport Hovav 2005, Melloni 2011, Ravid and Avidor 1998: and references therein). For example, the action noun of the CiCeC verb *pirsem* 'publish' is *pirsum* (CiCuC), which denotes both the action of publishing and the noun 'publication'.

Each verb pattern has a participle pattern that is used to indicate present tense of verbs. Participle patterns are polycategorical, as they are also used to denote nouns and adjectives. For example, CoCeC is the participle pattern that is related to the CaCaC verbal pattern. For example, *limes* denotes both the participle form of the verb *lamad* 'learn' and the agent noun 'learner'. The participle form *meratek* (meCaCeC) denotes both the participle form of the CiCeC verb *ritek* 'fascinate' and the adjective 'fascinating'. In addition, some words are formed in participle patterns and do not have verbal counterparts. For example, the agent noun *šoter* 'policeman' is formed in the CoCeC pattern and there is no CaCaC verb like **šatar*. Participle patterns are in

general multifunctional, as can be attested in other languages as well (for instance, verb-to-adjective conversion in French according to Tribout (2010)).

There is a special group of disyllabic patterns called Segolates: CéCeC, CáCaC, C6CeC and C6CaC. They differ from other patterns in three main aspects (see Bat-El 2012, Bolozky 1995, Schwarzwald 2002: among others). First, while most words that are formed in patterns have ultimate stress (e.g. *masrék* ‘a comb’) the Segolate patterns always have penultimate stress, e.g. CéCeC (*kéšer* and C6CeC (*tóxen* ‘content’).² Second, they are not associated with typical meanings and host a variety of nouns. Third, their inflectional paradigms exhibit three surface stems. For example, *kéšer* ‘relation’, *kšar-im* ‘relations’ and *kšr-i* ‘my relation’.

2.1.3. Root types and relations between roots

Most roots consist of 3 consonants, but four-consonant roots are also present in the lexicon in a non-negligible proportion. These are found almost exclusively in the CiCeC, CuCaC and hitCaCeC patterns whose prosodic structure can accommodate more than 3 consonants, e.g. p-rs-m for *pirsem* (CiCeC) ‘publish’. Some roots are weak in the sense that one or more of the consonants do not surface in all forms or do not surface at all. For example, the root of the verb *rac* (CaCaC) ‘run’ is r-w-c, where the w never surfaces and can only be associated with the verb through diachronic analysis. In addition, some phones undergo phonological alternation in the transition between patterns, e.g. stop-fricative alternations, as in *gavar* (CaCaC) ‘increase_{inchoative}’ - *higbir* (hiCCiC) ‘increase_{transitive}’.

By default, a relation connects two or more items that share the same root. This is one of the main features of Semitic morphology that is responsible for a rich system of derivational paradigms that revolve around the consonantal root. However, there are relations that connect items with different roots. These particular relations surface in cases where a consonant is added to the root. This type of relation creates a new family, and its members share the new root. The two families form different paradigms. Let us demonstrate it with respect to the pair *tadrix* ‘briefing_N’ - *tidrex* ‘debrief_V’. The taCCiC pattern, which includes the prefix *ta-*, is used for the formation of different nouns that can be related to verbs in different patterns, e.g., *hidpis* (hiCCiC) ‘print_V’ - *tadpis* ‘printout_N’. The noun *tadrix* ‘briefing’ is formed in the taCCiC pattern, and is semantically related to the hiCCiC verb *hidrix* ‘guide_V’ and the haCCaCa action noun *hadraxa* ‘guidance_N’. The three words form a derivational family sharing the consonantal root d-r-x. The verb *tidrex* ‘debrief_V’ is formed in the CiCeC pattern based on the noun *tadrix*, taking the t consonant of the derivational prefix *ta-* as part of the new root t-dr-x. The CiCeC pattern is paradigmatically connected to the CiCuC pattern of action nouns (*tidrux* ‘debriefing_N’) and to the passive CuCaC pattern (*tudrax*

²When words and patterns have penultimate stress, this is marked throughout the paper by an acute accent. Otherwise, patterns and words are left unmarked.

'be debriefed_V'). The CiCeC pattern of *tidrex* induces new types of relations within its new family.

2.2. Other word formation strategies

In contrast to verbs, the formation of nouns and adjectives is based on a variety of word formation strategies, most of which are highly productive in European languages. Nouns, for example, can be 'raw'³, (*daf* 'page'), borrowed (*lazanya* 'lasagna'), and can be formed via different word formation processes. Hebrew has a set of derivational affixes that are used for the formation of nouns and adjectives. Affixes can be attached to different stems with or without a morphological structure. They can be attached to raw stems. For example, the noun *yam* 'sea' takes the suffixes *-i* and *-ay* to derive the adjective *yam-i* 'marine' and the agent noun *yam-ay* 'sailor'. The adjective *kal* 'easy' takes the suffix *-ut* to derive the abstract noun *kal-ut* 'easiness'. Affixes are also attached to words with root and pattern. For example, the agent noun *naḡar* 'carpenter' is formed in the CaCaC pattern, and the suffix *-iya* is attached to form the location noun *naḡar-iya* 'carpentry shop'. Some words undergo morpho-phonological alternations when affixes are attached. For example, *šémeš* 'sun' undergoes two alternations in the formation of the adjective *šimš-i* 'sunny'; the first vowel changes from *e* to *i*, and the second vowel is deleted.

In addition to affixation, nouns and adjectives are formed by other word formation strategies like reduplication (*xatul* 'cat' - *xataltul* 'kitten'), acronym formation (e.g. *ramax* 'department chair', based on *roš* 'head' and *maxlaka* 'department'), blending (e.g. *midrexov* 'pedestrian mall', based on *midraxa* 'sidewalk' and *rexov* 'street') and compounding (e.g. *bet-sefer* 'school', lit. house-book).

2.3. Word-based approach to Semitic morphology

The design of *Démonette* (Section 3), and specifically its implementation into *Hebrewnet*, relies on word-based models to word formation. The word-based approach, originally proposed in (Aronoff 1976), assumes that the mental lexicon consists of actual words rather than morphemes, roots or coded concepts. Aronoff's main claim is that a word is formed by applying a Word Formation Rule (WFR) to an existing word or stem. They serve for producing and understanding new words, which may be added to the speaker's lexicon and as redundancy rules (Jackendoff 1975) defining morphological relations. Such a view assumes a phonological representation of words in the lexicon. The distinction between a root/morpheme-based morphology and a word-based morphology corresponds to the traditional distinction between 'item and arrangement' models and 'item and process' models respectively (Anderson 1992,

³we use 'raw' following Schwarzwald (2002), to indicate that a word has no complex morphological structure. It is not derived from another word, it is not formed in a pattern and does not consist of affixes.

Hockett 1954, Matthews 1972, 1974). In the former model, morphemes are the basic units of meaning and they are arranged linearly, while in the latter model, word structure is specified by a series of processes.

Semitic morphology raises questions about the exact processes that take place in word formation. We adopt the theory of Stem Modification (Bat-El 1994, 2017, 2019, McCarthy and Prince 1990, Steriade 1988), which accounts for generalizations about morpho-phonological alternations by allowing for stem-internal adjustments rather than positing the extraction of a consonantal root (Bat-El 1986, Davis and Zawaydeh 2001, Farwaneh 1990, Goldenberg 1994, Hoberman 1992, Idrissi and Kehayia 2004, McCarthy 1981, McCarthy and Prince 1986, Ornan 1983, Yip 1988: among others). Stem modification accounts for the transfer of information like prosodic structure from a base form to a derived form. It also provides a uniform account for morphological phenomena in non-Semitic languages, which are similar to those of non-concatenative morphology, e.g. ablaut in *sing/sang/song* (Bat-El 2002). Various studies have highlighted the absence of motivation for assuming an independent mechanism of root extraction (Aronoff 2007, Bat-El 1994, 2017, Benmamoun 2003, Bolozky 1999, 2012, Hammond 1988, Heath 1987, Kihm 2011, McCarthy and Prince 1990, Ratcliffe 1997, Rose 1998, Ussishkin 2005: among others). The status of the consonantal root is under an ongoing debate and there are different approaches with regard to its necessity and the actual mechanism that applies in word formation (Faust 2019, Nevins 2005, Rasin et al. 2021). It is important to emphasize that root-based approaches do not assume that Semitic word formation relies only on the representation of the consonantal root. Under such approaches, some words are derived directly from roots, while other words are derived directly from words (Arad 2005, Doron 2003, Faust 2015, Kastner 2019, 2020). Words that are derived from other words via non-concatenative morphology have to conform to one of the existing patterns. This is executed via “template imposition” (Faust and Hever 2010), where the pattern is imposed on the derived word based on its base. The question under debate is about the exact process that template/pattern imposition involves.

The design of Hebrewnet is based on a Stem Modification approach, as it represents, among other features, alternations that take place in the transition between words within paradigms. Such alternations relate both to the consonantal root and other parts of words. As will be detailed in Section 4, and demonstrated with respect to the cases studies in Section 5, the design of Hebrewnet provides the relevant information that is needed to examine structural relations between words which are formed in non-concatenative morphology (in addition to words formed by other processes), and such relations go beyond the consonantal root.

3. Démonette's principles

The founding principles of Démonette (Hathout and Namer 2016, Namer and Hathout 2020) that have been applied to Hebrewnette are the following (see also Laks and Namer 2020):

- Each entry describes a derivational relation between two lexemes, that is, unmarked words.
- Entries form derivational families represented by connected graphs, where derivational families are defined as sets of derivationally related lexemes (Hathout 2011).
- Lexemes and relations are described in two separate tables. The table of lexemes displays properties of words independent of the morphological relations these words can be involved in.
- Derivational relations occur for any pair of members of a given family. Relations are labelled according to their specific properties, as well as the properties (morphological, categorical, semantic, phonological) of the lexemes connected by the relation.
- These complex labels are the combination of several feature values. We exemplify them with the family of *banque* 'bank_N', *banquier* 'banker_N', *bancaire* 'of a bank_A', *interbancaire* 'between banks_A'.

Relations are distinguished according to their orientation, that is, whether one of the two connected lexemes is the ancestor of the other (in Table 1-a, *as2des* says that *banque* is the ancestor of *bancaire*; in Table 1-a', the reverse relation *des2as* indicates that the lexeme L1 *bancaire* is a descendant of L2 *banque*), or not (Table 1-b). Examples like Table 1-b include instances of cross-formation, where two co-derived words (like *prédateur* 'predator_{Nmas}' and *prédatrice* 'predator_{Nfem}', in French, Table 1-c) may lack a common ancestor (e.g. the verb **préder* is not attested). Orientation may be undecidable, and therefore labelled NA, as with the (*performant_A* 'performing', *performance_N* 'performance') conversion in Table 1-h.

When relevant, "ancestorhood" is evaluated based on semantic criteria. Let us examine the pair (*vivisecter_V* 'vivisect', *vivisection_N* 'vivisection'), Table 1-d. From a morphological (that is, formal) point of view, the noun seems to be derived from the verb by suffixation in *-ion* (as for example, *infection_N* 'infection' is derived from *infecter_V* 'infect'). However, the verb *vivisecter* is much more recent than *vivisection* and much less frequent (the Google search with the infinitive verb form results in approximately 2.000 hits, whereas it results in more than 2.5 millions with the singular noun form). Most importantly, unlike *infection_N* which is undoubtedly interpreted as the action noun of *infecter_V*, *vivisection_N* cannot be defined with respect to the semantic content of the verb (*vivisection* is by no mean 'the action of *vivisecter_V*'); on the contrary, the noun is the semantic base of the verb, which can be defined as 'to practice a *vivisection_N*'. The orientation value *as2des* indicates that the noun is the ancestor

	L1	L2	Orientation	Morpho. pattern1	Morpho. pattern2	Complexity
a	banque	bancaire	as2des	X	Xaire	simple
a'	bancaire	banque	des2as	Xaire	X	simple
b	banquier	bancaire	indirect	Xier	Xaire	simple
e	bancaire	interbancaire	as2des	X	interX	motiv-form
f	banque	interbancaire	as2des	X	interXaire	motiv-sem
g	banquier	interbancaire	indirect	Xier	interXaire	complex
c	prédateur	prédatrice	indirect	Xeur	Xrice	simple
d	vivisection	vivisecter	as2des	Xion	X	simple
h	performant	performance	NA	X	X	simple

Table 1. Orientation and complexity of derivational relations in *Démonette*

	L1	L2	Pattern1	Pattern2	Formal variation	Complexity	Cross-definition
a	fleur	fleurette	X	Xette	NA	simple	a <i>fleurette</i> is a small <i>fleur</i>
b	fleur	floral	X	Xal	/œ/ ~ /ɔ/	simple	—
c	fleur	anthophobe	X	Xphobe	NA	motiv-sem	they who fear <i>fleurs</i> are <i>anthophobes</i>

Table 2. Formal variation and cross-definitions in *Démonette*

in the relation. Divergence between form and meaning, such as the verb formation in Table 1-d corresponds to the phenomenon of so-called back-formation (Becker 1993).

What makes rows (d) and (a) (for instance) in Table 1 two different cases is the combination of the orientation value with the morphological pattern of the two lexemes involved. As we can see, X is the pattern of the descendant of the relation in Table 1-d, whereas it is that of the ancestor in Table 1-a. Notice that the pattern of a lexeme depends on the relation where it occurs. For instance, in (*bancaire, banque*) Table 1-a, the pattern of *bancaire* is $X\alpha i re$, and that of *banque* is X, where X represents the stem /bāk/ they share. In contrast, *bancaire* in Table 1-e is connected to *interbancaire*, and the shared stem is /bākεε/, therefore the pattern of *bancaire* is X and that of *interbancaire* is interX.

Another key feature of a derivational relation is its complexity. For regular relations, the value is simple when either one of the two lexemes is the base for the derivation of the other Table 1-a, or when both lexemes are daughters of the same base (even when this base is not or no more attested) as in Table 1-b,c. It is complex otherwise, as in Table 1-g. Not all derivational relations are morphologically canonical (in the sense of Corbett (2010)). They may involve form-meaning mismatches (Hathout and Namer 2014, Namer and Hathout 2020). This is the case for parasynthetic phenomena (Iacobini 2020), of which *interbancaire* is one of the many illustrations (Table 1-e,f). On the one hand, this adjective is formally derived from the adjective *bancaire* (Table 1-e). But on the other hand, its semantic content directly depends on that of the noun *banque*, since *interbancaire* means ‘between several banks’ and not ‘between things related to banks’. So, semantically, *interbancaire* is derived from *banque*. This dual motivation is expressed by two new values of complexity: *motiv-form* indicates that the relation is uniquely motivated formally (but not semantically), and *motiv-sem* expresses direct interpretative filiation (but a lack of formal transparency).

Besides structural properties, a relation within a derivational family carries phonological features that describe the way the relation affects the stems of the related lexemes, as illustrated in Table 2 with examples of lexemes derivationally connected to the noun *fleur* ‘flower’. There is formal identity when at least one of L1’s stems is identical to one of L2’s stems, as in Table 2-a. Otherwise, phonological variations are ranked according to morphophonological features. In Table 2-b, the only variation at play is an instance of vowel backness. For stems that are historically related but are unrelated from a synchronic perspective, there is no phonological variation encoded, but the value of complexity is *motiv-sem*. This is the case of *antho-* (/ãto/) in Table 2-c; this Greek learned suppletive component of the noun *fleur* /flœε/ ‘flower’ occurs almost only in neoclassical compounds.

Finally, relations are encoded with features describing their semantic properties. Based on the ontological class of each lexeme, these properties include the semantic category of the relation, eg., *agent-activity* for (*prédateur, prédation*), *location-agent*, for (*banque, banquier*), or *identity* for relations between words with the same semantic content e.g. (*banque, bancaire*) (in line with Spencer’ (2013) notion of transposi-

tion). In addition, relations are described by means of a paraphrase cross-defining the related words, cf. Table 2-a,c last column.

To sum up, the database is deliberately designed as highly redundant. Each lexical unit has as many derivational descriptions as it has multiple relations within its family. In addition to the properties of its relation with other words, each lexical unit is defined by features independent of the relations in which it is found (e.g. its inflectional paradigm, part of speech, ontological category, frequency).

A morphological description is therefore the result of the interaction of formal, categorical and semantic properties. These three levels of description are autonomous, which allows us to represent non-canonical phenomena, such as derivations involving meaning-form discrepancies (back- & cross-formation, parasynthesis, conversion, bracketing paradoxes, etc. see Hathout and Namer (2014)) straightforwardly. At each level of description, properties are abstracted away into patterns that generalize the different sorts of regularities that can be found in the constructed lexicon: phonological, semantic, morphological. In other words, *Démonette* implements the principles of the paradigmatic approach to morphology (Bonami and Strnadová 2018, Hathout and Namer 2022): (i) relations where the same lexeme is involved combine two-by-two and form connected graphs that represent derivational families; (ii) these families can be superposed when the relations between their members instantiate the same abstract properties. In sum, the architecture is designed to integrate paradigmatic relations in morphology, which is also a characteristic of Hebrew, as we have seen in Section 2.

4. The Hebrewnet database: basic description

The design of Hebrewnet, based on the basic principles of *Démonette*'s annotation system, makes its features suitable for capturing both morphological and semantic relations between Hebrew words, regardless of the type of morphology (i.e. concatenative or non-concatenative). Other morphological tools and resources for Hebrew and other Semitic languages exist: see, for example, Daya et al. (2008), Itai and Wintner (2008), Klimek et al. (2016), Neme (2011), Nir et al. (2013), Singh and Habash (2012), Wintner (2004). However, they rely mostly on the consonantal root as the central entity used as a base for word formation, which implies that family networks are oriented tree-shaped graphs, where only ancestor-descendant relations are represented, and not paradigmatic relations between words. The design of Hebrewnet relies on a word-based approach to morphology, and it therefore allows a separation between structural and semantic properties, in the analysis of such paradigmatic relations. Note, however, that the properties of Hebrewnet can be used to analyse Hebrew morphology under both word-based and root-based approaches.

Given the non-concatenative nature of the morphology of Semitic languages (Section 2) and the structures already present in *Démonette* (Section 3), a number of extensions are necessary when transposing *Démonette* for the analysis of derivation in

Hebrew. In a nutshell, these extensions regard the description of (relations between) patterns and (relations between) roots (see Section 2.1.3). Other typical features of Hebrew, e.g. those presented in Section 2.1.1, are in line with the predictability of lexical semantic properties (argument structure, agentivity) of words whose pattern belongs to the same derivational paradigm.

First, new attributes are needed to describe the **internal (morphological) structure** and the **root** of the lexemes connected by a derivational relation:

- We have seen (Section 2.1.1) that verbs have root and pattern, while nouns and adjectives can be formed both by root and pattern (Section 2.1.2) and by other word formation processes (Section 2.2). The pattern is indicated for words that are formed via non-concatenative morphology, e.g. CiCeC and CaCaC, Table 3-a,b. Some patterns include affixes, e.g. hiCCiC, Table 3-c, and hitCaCeC, Table 3-d. In addition, derivational affixes can be attached to words that have root and pattern CaCaC+iya, Table 3-e.
- Some words do not have root and pattern (Section 2.2). In that case, their morphological structure is coded as *raw*, Table 3-f, or *borrowed*, Table 3-g. The pattern of words that are formed in Segolate patterns (Section 2.1.2) are marked with an accent (Table 3-h,l).
- For words with a pattern, the morphological structure is displayed in the form of its vowel schema between “|” (|ie|, Table 3-a), completed when relevant with pattern affixes (hi| \emptyset e|, Table 3-c, where \emptyset indicates an empty vowel position, hit| æ |, Table 3-d). In case of affixation, affixes which are not part of the pattern are separated from the base structure by ‘+’ (|aa|+iya, Table 3-e). This feature is significant also for borrowed words (Table 3-c), when they are the source of new patterns, cf. Table 4-c.
- When relevant, the formal variation between a word and its pattern is explicitly indicated: for example, Table 3-i, *rac* is an instantiation of the CaCaC pattern, where the second consonant of the pattern is missing C2: \emptyset , as well as its second vowel V2:a~ \emptyset (see Section 2.1.3).
- Raw and borrowed words have no root (NA stands for ‘not applicable’, Table 3-f,g). For other words, roots are classified according to their type. Three-consonant roots are labelled r(egular), e.g. Table 3-a,b. In four-consonant roots (r=4) the middle consonant position is instantiated by a cluster made of the second and third consonants (p-rs-m, Table 3-j). Phonological or orthographical identity between regular roots is expressed by a specific value. The value r-hom is used in case of homonymy. For instance, Table 3-k, the root s-p-r is used in two semantically unrelated derivational families: one containing *sipur*_N ‘story’, and one *sapar*_N ‘barber’ and both containing the ambiguous verb *siper*, which denotes either ‘tell’ or ‘cut hair’ (Table 3-k). The value r-hoph indicates a case of homophony. For example, the consonant /k/ of the phonological root k-š-r corresponds to two different spellings: כ (כשר) for the family of *kōšer*_N ‘ability’ (Table 3-l) and ק (קשר) for that of *kēšer*_N ‘relation’.

	Lexeme	Morpho- logical structure	Root type	Root	Morpho- phonological structure	Pattern- to-word alternation
a	limed 'teach'	CiCeC	r	l-m-d	ie	
b	nagar 'carpenter'	CaCaC	r	n-g-r	aa	
c	hirxiv 'make wide'	hiCCiC	r	r-x-v	hi 0i	
d	hitraxev 'become wide'	hitCaCeC	r	r-x-v	hit ae	
e	nagariya 'carpentry shop'	CaCaC+iya	r	n-g-r	aa +iya	
f	yam 'sea'	raw	NA	NA	W	
g	spam 'spam'	borrowed	NA	NA	0a	
h	kéšer 'relation'	CéCeC	r	k-š-r	ée	
i	rac 'run'	CaCaC	r- C2=W	r-W-c	a	C2:0, V2:a~0
j	pirsem 'publish'	CiCeC	r-4	p-rs-m	ie	
k	siper 'tell / cut hair'	CiCeC	r-hom	s-p-r	ie	
l	kóšer 'ability'	CóCeC	r-hoph	k-š-r	óe	

Table 3. Word representation in Hebrewnette

	L1/L2	L1/L2 Formal relation	L1/L2 Phonological variation	Root1/Root2	Root1/Root2 Relation
a	kivec _V /mexuvac _A	CiCeC/meCuCaC	C1:k~x	k-v-c / k-v-c	=
b	tadrix _N /tidrex _V	taCCiC/CiCeC	–	d-r-x/t-dr-x	CCC / tCCC
c	spam _N /hisvim _V	CCaC/hiCCiC	–	NA/s-p-m	NA

Table 4. Phonological properties of the L1/L2 relations in Hebrewnette

In addition to the annotations above, Hebrewnette describes **phonological variations** possibly involved by derivational relation (see Section 2.1.3), in line with the theoretical principles adopted in its conception, as mentioned in Section 2.3:

- between connected words - and more generally, the word patterns: in Table 4-a, there is a stop-fricative /k/~x/ alternation affecting the first consonant of the pattern;
- between roots: a derivational relation may trigger the creation of a new branch in a derivational family, characterized by an additional consonant in the root shared by the members of this new sub-family. Such derived roots (e.g. t-dr-x, Table 4-b) are formed by prefixing the base root (e.g. d-r-x) with the consonant (e.g. t) of the prefix included in the pattern of the words with this base root (e.g. taCCiC); relations between roots are encoded according to the value of the new root element, when this is relevant (e.g. CCC/tCCC in Table 4-b, see Section 2.1.3);
- even when one of the two related words does not have a root (in Table 4-c), as in the rootless and patternless borrowed noun *spam*, cf. Table 3-f, it has an apparent phonological representation of the form CCaC, Table 4-c, consistent with the morpho-phonological representation |θa|, and containing the consonant cluster sp. This representation may be relevant and taken into consideration in the formation of a verb like *hisvim*, that has the apparent root s-p-m⁴, and looks like other native Hebrew verbs with root and pattern (e.g. *hidpis* ‘print’). The root s-p-m and specifically the sp cluster are represented in the derived verb (Bat-El 1994, Bolozky 1978).

Regarding the **syntactic and semantic properties** of Hebrew words connected by paradigmatic relations (see Section 2), they are represented in Hebrewnette by several additional features (with respect to the ontological annotation already present in Démonette, see Section 3), where the main one describes verb argument structure. As

⁴We use the term “apparent root” to indicate that the base for word formation is either a raw native Hebrew word or a loan word (Section 2.2), which has no consonantal root. Since the formation of verbs based on such words must involve root and pattern morphology, the newly derived verbs seems to have a root. We thank an anonymous reviewer for the clarification.

	L1		L2		L1 argument structure	L2 argument structure
a	lamad _V	'learn'	nilmad _V	'be learned'	XY	YX
b	limed _V	'teach'	lamad _V	'learn'	WXY	XY
c	lamad _V	'learn'	hitlamed _V	'train _{intrans} '	XY	X
d	limed _V	'teach'	limud _N	'teaching'	WXY	WXY
e	lamad _V	'learn'	lamid _A	'learnable'	XY	Y

Table 5. Morphologically predictable argument structure of predicates in Hebrewnette

shown in Table 5, argument structure is encoded by means of the variables *X*, *Y*, *W*, which represent arguments of the predicate: in a family the same variable systematically corresponds to the same thematic role of the argument. For example, Table 5 displays an excerpt of the morphological network realized around the CaCaC transitive agentive verb *lamad* 'learn'.

- Table 5-a describes the relation between a transitive active (*XY*) structure and the corresponding passive one (*YX*), where *X* stands for the agent, and *Y* for the patient. Table 5-b describes the relation between a transitive causative verb (*WXY*) and its active transitive (*XY*) counterpart: *W* is an additional argument that causes the event represented by *lamad*.
- Table 5-c describes a relation between a transitive active agentive predicate (*XY*) and the corresponding intransitive verb (*X*).
- Moreover, since argument structure prediction goes beyond the verbal network in Semitic languages, the identity relation can also be defined between the structure of an active verb and that of its action noun (eg, Table 5-d, *limud_N*, which inherits its argument structure from *limed_V*), and the patient argument of a transitive verb can be passed to its related able-like adjective expressing potentiality (eg Table 5-e, *lamid_A*, which inherits its external argument *Y* from the argument structure *XY* of *lamad_V*).

In addition to the features and values which were presented, Hebrew morphology requires two distinct attributes to provide a precise description of the **orientation** of a relation: one for the morphological orientation, the other for the semantic one. This double organization, more complex than what is encoded in *Démonette* (see Table 1), is necessary for an accurate representation of form-meaning mismatches, as will be shown in Section 5.

5. Case studies

We now turn examine two case studies that deal with different theoretical aspects of Hebrew morphology that have been addressed in previous papers. We use these case studies to demonstrate how the properties of Hebrewnette allow us to provide em-

pirical evidence in order to answer theoretical questions and shed light on the structural and semantic relations between words that are formed via non-concatenative morphology.

5.1. Faithfulness constraints and competing patterns

This case study examines doublet formation of Hebrew instrument nouns (hereafter INs). This is a case in which two INs that share the same meaning and consonantal root are constructed in two different patterns. Such variation is shown in Table 6.

L1	L2	Gloss	Root	Pattern1	Pattern2
masnen	mesanen	'filter' _N	s-n-n	maCCeC	meCaCeC
maghec	megahec	'iron' _N	g-h-c	maCCeC	meCaCeC
magresa(t)[-kerax]	gores[-kerax]	'[ice-]crusher' _N	g-r-s	maCCeCa	CoCeC
maxlec[-pkakim]	xolec[-pkakim]	'[cork-]screw' _N	x-l-c	maCCeC	CoCeC

Table 6. Morphological variation of Instrument nouns (INs)

As shown in Table 6, *masnen*, for example, is formed in maCCeC while *mesanen* is formed in meCaCeC, and both nouns share the consonants s-n-n and denote 'filter'. Nouns formed in the patterns of the Pattern1 column are typically considered the prescriptive forms, unlike nouns formed in the patterns of the Pattern2 column, which have become more frequent in current usage; speakers demonstrate a tendency to use the non-prescriptive forms to different degrees (Boložky 1999, 2003). Regardless of the issue of the normative forms, both words share the same meaning and can be used in the same semantic-syntactic context. Laks (2015) shows that such doublet formation and lack thereof can be predicted based on morphological and semantic criteria. In this study, we address the morphological aspect of doublet formation and show that the properties of Hebrewnette allow us to predict which doublet member is preferred.

We begin with some background on Hebrew INs formation. There are two main groups of INs patterns. The participle patterns CoCeC, meCaCeC, maCCiC are ambiguous as they also denote the present tense of verbs, as illustrated in Table 7. The form *sorek*, for example, corresponds both to the noun 'scanner' and the present form of the verb *sarak* 'scan'.⁵

⁵The participle patterns can also denote agent nouns, e.g. *moxer* 'seller', related to the verb *maxar* 'sell', and also adjectives, e.g. *madhim* 'amazing', related to the verb *hidhim* 'amaze' (Section 2.1.2). Faust (2011) shows that agent nouns and INs are formed independently, i.e. without a corresponding verb, only in the CoCeC pattern. That is, other participle patterns do not host such independent nouns without a verbal alternate in the relevant pattern.

	Verb Pattern	Example	IN / Participle Pattern	Example
a	CaCaC	sarak 'scan'	CoCeC	sorek 'scanner'
b	CiCeC	yibeš 'dry'	meCaCeC	meyabeš 'drier'
c	hiCCiC	hismix 'thicken' (liquids)	maCCiC	masmix 'thickener'

Table 7. INs formation in participle patterns

Other patterns that host INs are not used as verbs and are not related to a specific verbal pattern (hereafter 'non-participle patterns'). Some of these patterns are presented in Table 8. It is important to note that this is not an exhaustive list, but it represents the common patterns in which INs are formed. Some of them, e.g. maCCeC, are more typical for INs than others, e.g. CaCaC, and none of them is exclusively used for INs formation (see Bolozky 1999, Schwarzwald 2002: and references therein). For example, the noun *mirpéset* 'balcony' is formed in the miCCeCet pattern, but denotes a location rather than an IN. This corresponds to a different ontological value in Hebrewnette.

	Pattern	Example
a	maCCeC	maxbet _N 'bat'
b	maCCeCa	maclema _N 'camera'
c	miCCéCet	miklédet _N 'keyboard'
d	maCCéCet	madpéset _N 'printer'
e	CaCaC	vasat _N 'regulator'

Table 8. INs formation in non-participle patterns

INs doublets are formed in cases where an existing IN in a non-participle pattern takes an additional form in a participle pattern (Bolozky 2003, Laks 2015). This is shown in Table 9 where the form in (ii) is preferred over the form in (i) in both cases.

	Verb	Verb Pattern	Instrument noun	Nominal Pattern
a	sinen _V 'filter'	CiCeC	(i) masnen 'filter' (ii) mesanen	maCCeC maCaCeC
b	hidgiš _V 'emphasize'	hiCCiC	(i) madgeš 'marker' (ii) madgiš	maCCeC maCCiC

Table 9. Doublet formation of INs

One of the doublets is preferred due to faithfulness to the base from which it is derived. In Table 9-a, the formation of *mesanen* is more faithful to *sinen* as it involves only prefixation and changing one vowel, while the prosodic structure remains intact. In contrast, the formation of *masnen* changes the prosodic structure of the base, as it creates the consonant cluster *sn* that does not exist in the verb *sinen*. In Table 9-b, the formation of both *madgeš* and *madgiš* does not change the prosodic structure of the verb, but *madgiš* is more faithful because its second vowel /i/ is identical to the second vowel of the related verb *hidgiš*. The formation of both instrument nouns in (ii) involves fewer changes with respect to the verb, and as a result there is greater structural transparency between the verb and the instrument noun.

Hebrewnette provides the information required to compare concurrent INs according to their degree of faithfulness to their related verb, thanks to the value of the feature called `Morphophonological structure` we have introduced in Section 4, Table 3. As shown in Table 10, this feature allows computing the difference in the edit distance (known as Levenshtein distance) between the verb L1 and the ‘regular’ IN form L2 in Table 10-a,c on the one hand, and between L1 and the concurrent IN L2 in Table 10-b,d, on the other hand. The smaller the edit distance, the greater the faithfulness of L2 to L1, the more likely the L2 form.

We use a measure parametrized such that string modification is weighed according to the distance from the original syllabic and melodic structure. Therefore we decided that vowel substitution is twice as ‘cheaper’ (distance=1) as prefix insertion (or substitution or deletion) (distance=2). Moreover, it weights four times less than vowel insertion (or deletion) (distance=4), because the latter transformation involves “consonant (de)clusterization”, that is, either breaking consonant clusters that exist in the base, or creating consonant clusters that are not part of the base.

The IN in Table 10-b is preferred over the one in Table 10-a because its edit distance from the verbal base is 3, while in Table 10-a it is 6. Similarly, the IN doublet member in Table 10-d is preferred over the one in Table 10-c because its edit distance from the verb is smaller.⁶ As shown, the features encoded in Hebrewnette allow us to deduce predictions with respect to doublet formation and explain why one of the two doublets is preferred over the other.

Other theoretical hypotheses are empirically validated thanks to Hebrewnette annotations. This is what we show through a second case study.

5.2. Form/meaning mismatches

This second case study addresses transitivity alternations. Transitive-intransitive alternations within verbal systems have been the object of various studies including Alexiadou et al. (2006), Berman (1982, 1993), Borer (1991), Doron (2003), Haspel-

⁶Notice that both the Levenshtein measure used to predict the most likely IN competitor and the weight assigned to each criterion, in the two rightmost columns in Table 10 are external from the design of Hebrewnette and can be adjusted according to the need of the database users.

	L1	L2	Morpho-phonological Structure		L1 / L2 changes	Edit distance
			L1	L2		
a	sinen _V	masnen _N	ie	ma 0e	prefix insertion: ma- vowel insertion: i	6
b	sinen _V	mesanen _N	ie	me ae	prefix insertion: me- vowel change: i ↔ a	3
c	hidgiš _V	madgeš _N	hi 0i	ma 0e	prefix change: hi- ↔ ma- vowel change: i ↔ e	3
d	hidgiš _V	madgiš _N	hi 0i	ma 0i	prefix change: hi- ↔ ma-	2

Table 10. Predicting the outcome of INs competition in Hebrewnette

math (1987), Horvath and Siloni (2008, 2010), Koontz-Garboden and Levin (2005), Levin and Rappaport Hovav (1995), Pylkkänen (2008), Reinhart (1996), Rappaport Hovav and Levin (2007, 2012), Williams (1981). It is commonly assumed that different thematic realizations of the same concept are not accidental and that there are some sort of derivational relations between verbs that participate in such alternations.

These alternations have been addressed by syntactic, semantic, and morphological theories, attempting to shed light on both the morphological and the syntactic and semantic-thematic characteristics of such derivations

Causative / inchoative alternations can involve apparent morpho-semantic mismatches, as discussed in Borer (1991), Doron (2003), Haspelmath (1987, 1993), Horvath and Siloni (2010), Rappaport Hovav and Levin (2012) among many others, where semantic and morphological directions seem to collide. In each pair of verbs in Table 11, the semantic relation is similar, where the transitive verbs denote causation of change in Y's mental state, and the intransitive verbs denote the change in the mental state that Y undergoes (see Table 5 for the way semantic roles are assigned to the symbols X, Y etc.). However, their structural (morphological) relations are different. In Table 11-a, the morphological relation is formally oriented from the transitive verb to the intransitive verb, as the former is formed in an affixless pattern (CiCeC), while the latter is formed in a pattern with a prefix (hitCaCeC). In contrast, in the relation of Table 11-b, the transitive verb, formed in a pattern with a prefix (hiCCiC), is formally more complex than the intransitive one formed in CaCaC.

To represent the form-meaning mismatch illustrated in Table 11, Hebrewnette encodes separately semantic and structural information about the direction of derivational relations. Based on the orientation attribute and its 'as2des' and 'des2as' values, presented for Démonette in Section 3 (see Table 1), orientation is duplicated into

two properties: a formal and a semantic one. This organization enables an accurate representation of mismatches like with the two verb pairs of Table 11.

This is illustrated in Table 12. Since there are different approaches regarding semantic directionality for causative/inchoative alternations (see for example, the discussion in (Horvath and Siloni 2010)), we decided to encode semantic orientation as unspecified (NA) for both verb pairs, while the semantic difference between the transitive and the intransitive verb of each pair is given by the respective value of the argument-structure (in the Argument Structure columns) we introduced in Table 5.

In contrast, Formal (or morphological) orientation is determined by the presence of a prefix (and lack thereof) in one of the verbs in each pair. For the first verb pair, the Formal Orientation goes from L1 to L2 (as2des) because only L2 (*hityaʔeš*) consists of a prefix (*hit-*), while the opposite orientation holds for the second verb pair, because only L1 (*hitsis*) consists of a prefix (*hi-*). The value of the morphological orientation can be automatically computed using the same edit distance principles used for predicting the outcome of the competition between IN patterns (Section 5.1, Table 10), as shown in the last column of Table 12.

Transitive V		Pattern	Intransitive V		Pattern
yiʔeš	'X make Y desperate'	CiCeC	hityaʔeš	'Y become desperate'	hitCaCeC
hitsis	'X make Y agitated'	hiCCiC	tasas	'Y become agitated'	CaCaC

Table 11. Transitive/intransitive alternation

L1	L2	Argument structure		SO	Morpho-phonological structure		FO	L1 / L2 Changes
		L1	L2		L1	L2		
yiʔeš	hityaʔeš	XY	Y	NA	ie	hit ae	as2des	prefix insertion: hit- vowel change: i ↔ a prefix insertion: hi- vowel insertion: a vowel change: i ↔ a
hitsis	tasas	XY	Y	NA	hi 0i	aa	des2as	

Table 12. Formal (FO) vs. semantic (SO) orientation in Hebrewnette

6. The Hebrewnet prototype

The above pilot study has resulted in the design of a prototypical database for Hebrew. The goal is to cover all the morphological phenomena that the complex lexicon of Hebrew may present, and more generally, all the paradigmatic aspects related to non-concatenative morphology. The collection of relations to be included in this prototypical version of the database combines two strategies.

- The first one is entirely manual; it consists in selecting a set of families according to the property(ies) that distinguish each of them, in order to test the expressive power of the notation system presented in Section 4.
- The second one is semi-automatic, and takes advantage of the partial predictability of verbal paradigms. It includes the automatic generation (followed by a manual post-editing) of families centered on pivot verbs instantiating the CiCeC pattern.

The first strategy follows “exemplar-based” principles. The word families that were selected intended to cover typical relations between Hebrew words. Therefore, the whole set of properties described in Sections 2, 4, and 5 correspond to at least one word-family included in the database, and include, among others:

- the type of root (Section 2.1.3 and Table 3),
- the mode of lexeme formation: relations between patterns (Section 2.1.1 and 2.1.2) and affixation (Section 2.2),
- the phonological alternation between a pattern and the form which realizes it (last column of Table 3),
- the pattern-to-pattern phonological variations (Table 4),
- the formation of subfamilies by root-to-root relations (Section 2.1.3 and Table 4),
- the form-meaning discrepancies (Section 5.2),
- the different cases of argument structure (Table 5).

This “family-centered” coverage describes the relations among a total of 245 lexemes belonging to 28 different families of different size (containing at least 4 members, as in the family of *yam*_N ‘sea’, and at most 28, as in the family of *kešer*_N ‘relation’).

The second strategy, of “pattern-centered” coverage, is based on the regularities observed empirically in families based on CiCeC verbs. CiCeC has been selected because it is the most productive pattern for newly coined verbs and subsequent families, with a relatively higher predictability. In contrast, other verbal patterns (CaCaC, niCCaC) are either unproductive (few new families come from verbs in the CaCaC pattern) or unpredictable: apart from the active-passive relation, it is more difficult to predict the content and size for families of verbs in the other patterns, that is, hiCCiC or hitCaCeC. Regularities regarding CiCeC verbs have been encoded in order to semi-automatically generate and annotate derivational families based on an initial list of 10 CiCeC verbs. Relying on the fundamentally paradigmatic nature of the Hebrew verbal lexicon the following predictions have been implemented:

- CiCeC verbs are likely to realize active, transitive, dynamic predicates, e.g. *xibek* ‘hug_V’, *kivec* ‘shrink_V’, *nihel* ‘manage_V’;
- they are related to a CiCuC action noun (*xibuk* ‘hug_N’, *nihul* ‘management_N’), a resultative adjective in the meCuCaC participle pattern (*menohal*⁷ ‘managed_A’). CiCeC is also derivationally related to the meCaCeC participle pattern that can surface as an agent noun (*menahel* ‘manager_N’) an instrument noun, or an adjective (Section 2.1.2);
- when it is attested, their hitCaCeC related verb is intransitive, typically inchoative (*hitkavec* ‘become shrunk_V’), reflexive (*hitraxec* ‘wash oneself_V’) or reciprocal (*hitxabek*, ‘hug each other_V’).

From these 10 CiCeC verbs, the program produced 70 new annotated lexemes (after manual verification, 20 of them are discarded): each CiCeC verb is the source of a family of 6 members on average. As each member in a family is linked to all the others, this amounts to supplementing the 245 initial word pairs with 90 new fully documented entries.

7. Conclusions

This paper presented the main principles of the architecture of Hebrewnette, a derivational database for Hebrew, and its properties. We accounted for the adaptations that were made on the Démonette database, which was originally designed for Romance Morphology. Focus was on non-concatenative formation, which is highly typical of Hebrew and Semitic languages in general. We outlined the way words were coded with respect to their root and pattern. Taking a word-based approach for word formation, Hebrewnette is based on coding relations between words, and specifically for Hebrew, relations between roots and patterns. It is based on separate descriptions of semantic and structural relations so that each type of relation can be examined according to different criteria, e.g. direction of derivation and morphophonological alternations (if any). The features and feature values in the Hebrewnette database intertwine with the content of Démonette, to account for the properties of non-concatenative morphology. However, these additions do not compromise the architecture of Démonette; the global structures of the two databases are superimposable, which allows us to envisage a total interoperability between the two systems (and more generally between the morphologies of Romance languages and Semitic languages). We examined two cases that demonstrate how generalizations about the nature of Hebrew morphology can be captured based on the properties of Hebrewnette. The case of doublet formations of instrument nouns demonstrates the importance of the Hebrewnette representation of structural relations between words and how such representations can provide predictions with respect to the likelihood of

⁷The /u/ to /o/ variation between the pattern meCuCaC and the word *menohal* is due to the fact that the second consonant of the root /h/ is a glottal.

doublets to be derived. The case of form/meaning mismatches in transitivity alternations sheds light on the importance of distinguishing between semantic and structural descriptions of relations between words and the relevant implementation of describing relations between words that are formed in patterns.

Acknowledgements

This research has been realized as part of the project Demonext⁸, supported by the Agence Nationale de la Recherche, grant number: ANR-17-CE23-0005. It has been also supported by the Chateaubriand Fellowship Program of the French Embassy in Israel.

Bibliography

- Alexiadou, Artemis. *Functional structure in nominals: nominalization, and ergativity*. John Benjamins, Amsterdam, 2001. doi: 10.1075/la.42.
- Alexiadou, Artemis, Elena Anagnostopoulou, and Florian Schäfer. The properties of anticausatives crosslinguistically. In Frascarelli, Mara, editor, *Phases of Interpretation*, pages 187–212. Mouton, Berlin, 2006. doi: 10.1515/9783110197723.4.187.
- Anderson, Stephen R. *A-Morphous Morphology*. Cambridge University Press, Cambridge, UK, 1992. doi: 10.1017/CBO9780511586262.
- Arad, Maya. *Roots and Patterns: Hebrew morpho-syntax*. Springer, Dordrecht, 2005.
- Aronoff, Mark. *Word Formation in Generative Grammar*. Linguistic Inquiry Monographs. MIT Press, Cambridge, MA, 1976.
- Aronoff, Mark. *Morphology by Itself. Stems and Inflectional Classes*. MIT Press, Cambridge, MA, 1994.
- Aronoff, Mark. In the Beginning was the Word. *Language*, 83:803–830, 2007. doi: 10.1353/lan.2008.0042.
- Bat-El, Outi. *Extraction in Modern Hebrew morphology*. Phd thesis, UCLA, California, 1986.
- Bat-El, Outi. Stem modification and cluster transfer in Modern Hebrew. *Natural Language and Linguistic Theory*, 12:572–596, 1994. doi: 10.1007/BF00992928.
- Bat-El, Outi. Semitic verb structure within a universal perspective. In Shimron, Joseph, editor, *Languages processing and acquisition in languages of semitic, root-based, morphology*, pages 29–59. Benjamins, Amsterdam, 2002. doi: 10.1075/lald.28.02bat.
- Bat-El, Outi. Prosodic alternations in Modern Hebrew segolates. In Muchnik, Malka and Zvi Sadan, editors, *Studies on Modern Hebrew and Jewish Languages*, pages 116–129. Carmel Press, Jerusalem, 2012.
- Bat-El, Outi. Word-based items-and processes (WoBIP): Evidence from Hebrew morphology. In Bower, Claire, Laurence Horn, and Raffaella Zanuttini, editors, *On Looking into Words (and beyond)*, pages 115–135. Language Science Press, Berlin, 2017.

⁸<https://www.demonext.xyz/>

- Bat-El, Outi. Templatic morphology (clippings, root-and-pattern). In *Oxfors Research Encyclopedia of Linguistics*. Oxford University Press, Oxford, 2019.
- Batsuren, Khuyagbaatar, Gábor Bella, and Fausto Giunchiglia. MorphyNet: a Large Multilingual Database of Derivational and Inflectional Morphology. In *Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 39–48, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.sigmorphon-1.5. URL <https://aclanthology.org/2021.sigmorphon-1.5>.
- Becker, Thomas. Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. *Yearbook of Morphology 1992*, pages 1–25, 1993. doi: 10.1007/978-94-017-3712-8_1.
- Benmamoun, Elabbas. The role of the imperfective template in Arabic morphology. In Shimron, Joseph, editor, *Language Processing and Acquisition in languages of Semitic, root-Based, morphology*, pages 99–114. John Benjamins, Amsterdam, 2003. doi: 10.1075/lald.28.05ben.
- Berman, Ruth. Modern Hebrew structure. Report, University Publishing Projects, 1978.
- Berman, Ruth. Verb-pattern alternation: the interface of morphology, syntax, and semantics in Hebrew child language. *Journal of Child Language*, 9:169–191, 1982. doi: 10.1017/S030500090000369X.
- Berman, Ruth. Marking of verb transitivity by Hebrew-speaking children. *Journal of Child Language*, 20:641–669, 1993. doi: 10.1017/S0305000900008527.
- Berman, Ruth and Batia Seroussi. Derived nouns in Hebrew: Structure, meaning, and psycholinguistic perspectives. *Italian Journal of Linguistics*, 23(1):105–125, 2011.
- Bolozky, Shmuel. Word formation strategies in MH verb system: denominative verbs. *Afroasiatic Linguistics*, 5:1–26, 1978.
- Bolozky, Shmuel. The segolates: Linear or discontinuous derivation? In Schwarzwald, Ora R. and Izchak M. Schlesinger, editors, *Hadassah Kantor Jubilee Book (in Hebrew)*. Bar-Ilan University, Ramat Gan, 1995.
- Bolozky, Shmuel. *Measuring productivity in word formation: the case of Israeli Hebrew*. Brill, Leiden, 1999. doi: 10.1163/9789004348431.
- Bolozky, Shmuel. Phonological and morphological variations in spoken Hebrew. In Hary, Benjamin H., editor, *Corpus Linguistics and Modern Hebrew*, pages 119–156. Rosenberg School of Jewish Studies, Tel Aviv, 2003.
- Bolozky, Shmuel. More on linear vs. discontinuous derivation in Israeli Hebrew morphology. In Muchnik, Malka and Zvi Sadan, editors, *Studies in Modern Hebrew and Jewish*, pages 50–59. Carmel, Jerusalem, 2012.
- Bonami, Olivier and Jana Strnadová. Paradigm structure and predictability in derivational morphology. *Morphology*, pages 1–31, 2018. doi: 10.1007/s11525-018-9322-6.
- Borer, Hagit. The causative-inchoative alternation: a case study in parallel morphology. *The Linguistic Review*, 8:119–158, 1991. doi: 10.1515/tlir.1991.8.2-4.119.
- Borer, Hagit. Derived Nominals and the Domain of Content. *Lingua*, 141:71–96, 2014. doi: 10.1016/j.lingua.2013.10.007.

- Comrie, Bernard and Sandra A. Thompson. Lexical Nominalization. In Shopen, Timothy, editor, *Language Typology and Syntactic Description*, volume III, pages 334–381. Cambridge University Press, Cambridge, 2007. doi: 10.1017/CBO9780511618437.006.
- Corbett, Greville G. Canonical derivational morphology. *Word Structure*, 3(2):141–155, 2010. doi: 10.3366/word.2010.0002.
- Davis, Stuart and Bushra Adnan Zawaydeh. Arabic hypocoristics and the status of the consonantal Root. *Linguistic Inquiry*, 32(3):512–520, 2001. doi: 10.1162/002438901750372540.
- Daya, Ezra, Dan Roth, and Shuly Wintner. Identifying Semitic Roots: Machine Learning with Linguistic Constraints. *Computational Linguistics*, 34(3):429–448, 2008. doi: 10.1162/coli.2008.07-002-R1-06-30.
- Doron, Edit. Agency and voice: The semantics of the Semitic templates. *Natural Language Semantics*, 11:1–67, 2003.
- Farwaneh, Samira. Well-formed associations in Arabic: Rule or condition. In Eid, Mushira and John McCarthy, editors, *Perspectives on Arabic Linguistics*, volume II, pages 120–142. John Benjamins, Amsterdam/ Philadelphia, 1990. doi: 10.1075/cilt.72.08far.
- Faust, Noam. *La Morpho-Syntaxe Nominale de l'hebreu moderne du point de vue de la forme phonologique*. Thèse de doctorat, Université Paris-Diderot, 2011.
- Faust, Noam. A novel, combined approach to Semitic word-formation. *Journal of Semitic Studies*, LX(2):287–316, 2015. doi: 10.1093/jss/fgv001.
- Faust, Noam. New reasons to root for the Semitic root from Mehri and Neo-Aramaic. *The Linguistic Review*, 36(3):575–599, 2019. doi: 10.1515/tlr-2019-2030.
- Faust, Noam and Yarar Hever. Empirical and Theoretical Arguments in Favor of the Discontinuous Root in Semitic Languages. *Brill's Journal of Afroasiatic Languages and Linguistics*, 2: 80–118, 2010. doi: 10.1163/187666310X12688137960704.
- Goldenberg, Gideon. Principles of Semitic word-structure. In Goldenberg, Gideon and Schlomo Raz, editors, *Semitic and Cushitic Studies*, pages 10–45. Harrassowitz Verlag, Wiesbaden, 1994.
- Grimshaw, Jane. *Argument Structure*. MIT Press, Cambridge (MA) / London, 1990.
- Hammond, Michael. Templatic transfer in Arabic broken plurals. *Natural Language and Linguistic Theory*, 6:247–270, 1988. doi: 10.1007/BF00134231.
- Haspelmath, Martin. Transitivity alternations of the anticausative type. In *Institut für Linguistik (Köln). Abteilung Allgemeine Sprachwissenschaft: Arbeitspapier ; N.F., Nr. 5*. 1987.
- Haspelmath, Martin. *More on the typology of inchoative / causative verb alternations*, pages 87–111. John Benjamins, Amsterdam/Philadelphia, 1993. doi: 10.1075/slcs.23.05has.
- Hathout, Nabil. Une approche topologique de la construction des mots : propositions théoriques et application à la préfixation en *anti-*. In *Des unités morphologiques au lexique*, pages 251–318. Hermès Science-Lavoisier, Paris, 2011.
- Hathout, Nabil and Fiammetta Namer. Discrepancy between form and meaning in Word Formation: the case of over- and under-marking in French. In Rainer, Franz, Wolfgang U. Dressler, Francesco Gardani, and Hans Christian Luschützky, editors, *Morphology and meaning*, pages 177–190. John Benjamins, Amsterdam, 2014. doi: 10.1075/cilt.327.12hat.

- Hathout, Nabil and Fiammetta Namer. Giving Lexical Resources a Second Life: Démonette, a Multi-sourced Morpho-semantic Network for French. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1084–1091, 2016.
- Hathout, Nabil and Fiammetta Namer. ParaDis: a family and paradigm model. *Morphology*, 2022. doi: 10.1007/s11525-021-09390-w. URL <https://doi.org/10.1007/s11525-021-09390-w>.
- Hazout, Ilan. Action nominalizations and the lexicalist hypothesis. *Natural Language and Linguistic Theory*, 13:355–404, 1995. doi: 10.1007/BF00992736.
- Heath, Jeffrey. *Ablaut and ambiguity: Phonology of a Moroccan Arabic dialect*. State University of New York Press, Albany, 1987.
- Hoberman, Robert D. Local spreading. *Journal of Afroasiatic Linguistics*, 3:226–254, 1992.
- Hockett, Charles Francis. Two models of linguistic descriptions. *Words*, 10:210–234, 1954.
- Horvath, Julia and Tal Siloni. *Active lexicon: Adjectival and verbal passives*, pages 105–134. John Benjamins, Amsterdam/ Philadelphia, 2008. doi: 10.1075/la.134.05act.
- Horvath, Julia and Tal Siloni. *Lexicon versus syntax: Evidence from Morphological Causatives*, pages 153–176. Oxford University Press, Oxford, 2010.
- Iacobini, Claudio. Parasynthesis in Morphology. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, 2020. doi: 10.1093/acrefore/9780199384655.013.509.
- Idrissi, Ali and Eva Kehayia. Morphological units in the Arabic mental lexicon: Evidence from an individual with deep dyslexia. *Brain and Language*, 90:183–197, 2004. doi: 10.1016/S0093-934X(03)00431-0.
- Itai, Alon and Shuly Wintner. Language resources for Hebrew. *Language Resources and Evaluation*, 42:75–98, 2008. doi: 10.1007/s10579-007-9050-8.
- Jackendoff, Ray. Morphological and semantic regularities in the lexicon. *Language*, 51(3):639–671, 1975. doi: 10.2307/412891.
- Kastner, Itamar. Templatic morphology as an emergent property: Roots and functional heads in Hebrew. *Natural Language and Linguistic Theory*, 37(2):571–619, 2019. doi: 10.1007/s11049-018-9419-y.
- Kastner, Itamar. *Voice at the interfaces: The syntax, semantics and morphology of the Hebrew verb*. Language Science Press, Berlin, 2020. doi: 10.5281/zenodo.3865067.
- Kihm, Alain. Plural formation in Nubi and Arabic: A comparative study and a word-based approach. *Brill's Journal of Afroasiatic Languages and Linguistics*, 3:1–21, 2011. doi: 10.1163/187666311X562431.
- Kirov, Christo, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sebastian Mielke, Arya D. McCarthy, Sandra Kübler, David Yarowsky, Jason Eisner, and Mans Hulden. UniMorph 2.0: Universal Morphology. In Calzolari, Nicoletta, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7–12, 2018. European Language Resources Association (ELRA). ISBN 979-10-95546-00-9.

- Klimek, Bettina, Natanel Arndt, Sebastian Krause, and Timotheus Arndt. Creating Linked Data Morphological Language Resources with MMoOn. The Hebrew Morpheme Inventory. In Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 892–899. European Language Resources Association (ELRA), 2016.
- Koontz-Garboden, Andrew and Beth Levin. *The morphological typology of change of state event encoding*, pages 185–194. Università degli Studi di Bologna, Bologna, 2005.
- Kyjánek, Lukáš. Morphological Resources of Derivational Word-Formation Relations. Technical Report 61, ÚFAL - Charles University, Prague, 2018.
- Kyjánek, Lukáš, Zdeněk Žabokrtský, Magda Ševčíková, and Jonáš Vidra. Universal Derivations 1.0, A Growing Collection of Harmonised Word-Formation Resources. *The Prague Bulletin of Mathematical Linguistics*, 115:5–30, October 2020. ISSN 0032-6585. doi: 10.14712/00326585.003. URL <https://ufal.mff.cuni.cz/pbml/115/art-kyjanek-et-al.pdf>.
- Laks, Lior. Variation and change in instrument noun formation in Hebrew and its relation to the verbal system. *Word Structure*, 8(1):1–28, 2015. doi: 10.3366/word.2015.0071.
- Laks, Lior and Fiammetta Namer. Designing a derivational resource for non-concatenative Morphology: the Hebrewnet database. In Namer, Fiammetta, Nabil Hathout, Stéphanie Lignon, Zdeněk Žabokrtský, and Magda Ševčíková, editors, *DeriMo2021 - Third International Workshop on Resources and Tools for Derivational Morphology.*, pages 95–105. ATILF, 2020.
- Levin, Beth and Malka Rappaport Hovav. *Unaccusativity*. MIT Press, Cambridge, MA, 1995.
- Levin, Beth and Malka Rappaport Hovav. *Argument Realization*. Cambridge University Press, Cambridge, 2005. doi: 10.1017/CBO9780511610479.
- Matthews, Peter H. *Inflectional Morphology*. Cambridge University Press, Cambridge, 1972.
- Matthews, Peter Hugoe. *Morphology: an introduction to the theory of word-structure*. Cambridge University Press, London, 1974.
- McCarthy, Arya D., Miikka Silfverberg, Ryan Cotterell, Mans Hulden, and David Yarowsky. Marrying Universal Dependencies and Universal Morphology. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 91–101, Brussels, Belgium, November 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-6011. URL <https://aclanthology.org/W18-6011>.
- McCarthy, John. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, 12: 373–418, 1981.
- McCarthy, John and Alan Prince. *Prosodic morphology*. Thesis, 1986.
- McCarthy, John and Alan Prince. Foot and word in Prosodic Morphology: The Arabic broken plural. *Natural Language and Linguistic Theory*, 8:209–283, 1990. doi: 10.1007/BF00208524.
- Melloni, Chiara. *Event and Result Nominals: A Morpho-semantic Approach*. Peter Lang, Bern, 11 2011. ISBN 978-3-0343-0658-4.
- Namer, Fiammetta and Nabil Hathout. ParaDis and Démonette –From Theory to Resources for Derivational Paradigms. *The Prague Bulletin of Mathematical Linguistics*, 114:5–33, 2020. doi: 10.14712/00326585.001.

- Neme, Alexis Amid. A lexicon of Arabic verbs constructed on the basis of Semitic taxonomy and using finite-state transducers. In Sagot, Benoît, editor, *WoLeR 2011*, pages 79–86, 2011. URL [halshs-01186723](https://halshs.archives-ouvertes.fr/halshs-01186723).
- Nevins, Andrew. Overwriting does not optimize in non-concatenative word formation. *Linguistic Inquiry*, 36(2):275–287, 2005.
- Nir, Bracha, Brian MacWhinney, and Shuly Wintner. The Hebrew CHILDES corpus: transcription and morphological analysis. *Language Resources and Evaluation*, 47(4):973–1005, 2013.
- Ornan, Uzzi. *How is the Hebrew word formed?*, pages 13–42. Magnes, Jerusalem, 1983.
- Pylkkänen, Liina. *Introducing arguments*. MIT Press, Cambridge, 2008. doi: 10.7551/mitpress/9780262162548.001.0001.
- Rappaport Hovav, Malka and Beth Levin. *Deconstructing thematic hierarchies*, pages 385–402. CSLI Publications, Stanford, 2007.
- Rappaport Hovav, Malka and Beth Levin. *Lexicon uniformity and the causative alternation*, pages 150–176. University Press, Oxford, 2012.
- Rasin, Ezer, Omer Preminger, and David Pesetsky. A re-evaluation of Arad’s Argument for Roots. In *Proceedings of the 39th West Coast Conference on Formal Linguistics*, 2021. URL <https://ling.auf.net/lingbuzz/006077>.
- Ratcliffe, Robert R. *Prosodic templates in a word-based morphological analysis of Arabic*, pages 147–171. John Benjamins, Amsterdam/Philadelphia, 1997. doi: 10.1075/cilt.153.10rat.
- Ravid, Dorit. Internal structure constraints on new-word formation devices in Modern Hebrew. *Folia Linguistica*, 24:289–347, 1990. doi: 10.1515/flin.1990.24.3-4.289.
- Ravid, Dorit and Avraham Avidor. Acquisition of derived nominals in Hebrew: developmental and linguistic principles. *Journal of Child Language*, 25:229–266, 1998. doi: 10.1017/S0305000998003419.
- Ravid, Dorit, Orit Ashkenazi, Ronit Levie, Galit Ben Zadok, Tehila Grunwald, Ron Bratslavsky, and Steven Gillis. Foundation of the Early Root Category: Analyses of Linguistic Input to Hebrew-Speaking Children. In Berman, Ruth, editor, *Acquisition and Development of Hebrew: From Infancy to Adolescence*, pages 95–134. John Benjamins, Amsterdam, 2016.
- Reinhart, Tanya. Syntactic effects of lexical operations: Reflexives and unaccusatives. Report, University of Utrecht, 1996.
- Rose, Sharon. Triple take: Tigre and the case of internal reduplication. In Hayward, Hay, Jamal Ouhalla, and Denise Perett, editors, *Studies in Afroasiatic grammar*. John Benjamins, Amsterdam, 1998.
- Schwarzwald, Ora R. *Grammar and Reality in the Hebrew Verb*. Bar Ilan University Press, Ramat Gan, 1981.
- Schwarzwald, Ora R. *Studies in Hebrew Morphology (in Hebrew)*. The Open University, Tel Aviv, 2002.
- Singh, Nimesh and Nizar Habash. Hebrew Morphological Preprocessing for Statistical Machine Translation. In Cettolo, Mauro, Marcello Federico, Lucia Specia, and Andy Way, editors, *Proceedings of the 16th EAMT Conference*, pages 43–50. European Association for Machine Translation, 2012.

- Spencer, Andrew. *Lexical relatedness*. Oxford University Press, Oxford, 2013. doi: 10.1093/acprof:oso/9780199679928.003.0003.
- Steriade, Donca. Reduplication and transfer in Sanskrit and elsewhere. *Phonology*, 5(1):73–155, 1988.
- Tribout, Delphine. *Les conversions de nom à verbe et de verbe à nom en français*. Thèse de doctorat, Université Paris 7, 2010.
- Ussishkin, Adam. A fixed prosodic theory of nonconcatenative templatic morphology. *Natural Language and Linguistic Theory*, 23:169–218, 2005.
- Williams, Edwin. Argument structure and morphology. *The Linguistic Review*, 1:81–114, 1981. doi: 10.1515/tlir.1981.1.1.81.
- Wintner, Shuly. Hebrew computational linguistics: Past and future. *Artificial Intelligence Review*, 21(2):113–138, 2004. doi: 10.1023/B:AIRE.0000020865.73561.bc.
- Yip, Moira. Template morphology and the direction of association. *Natural Language and Linguistic Theory*, 6:551–577, 1988. doi: 10.1007/BF00134493.

Address for correspondence:

Fiammetta Namer
fiammetta.namer@univ-lorraine.fr
Université de Lorraine, CLSH
UFR SHS, dépt SDL
23 bd Albert 1er, BP 60446
54001 NANCY CEDEX, France