

Chapter 5

Multiword expressions and the Law of Exceptions

Koenraad Kuiper

University of Canterbury, New Zealand

This chapter proposes the existence of a linguistic universal, the Law of Exceptions. It hypothesizes that a relationship exists between the grammar of a language and its lexicon such that all regularities expressed in the grammar of a language are matched by exceptions which are manifested in the lexicon of that language. It is also proposed that lexical idiosyncrasies are of two types. Type 1 idiosyncrasies are in the nature of arbitrary restrictions on options provided in the grammar while Type 2 idiosyncrasies involve breaches of the rules of the grammar. To test this law requires an initial examination of the linguistic domains where it might be tested. As a preliminary step to testing these ideas, this chapter is a scoping exercise looking chiefly at the structural properties of a subset of multiword expressions (MWE). It shows, following Barkema (1996), that many properties of MWEs cross-classify. The aim of the overview is then to examine domains of the morphosyntax of any language which might be analysed for sources of structural idiosyncrasy and thus to determine how individual languages might vary in this respect. Languages of exemplification are English, which has a relatively fixed word order and slight inflectional system, and, to a lesser extent Dutch and Māori, an Oceanic language.

1 Introduction

In the traditional grammar-lexicon model of human linguistic knowledge, the grammar accounts for the regularities in the language which a native speaker is taken to have acquired and thus the predictabilities in its sentences. The lexicon has traditionally been ‘an appendix of the grammar, a list of basic irregularities,’ (Bloomfield 1933: 274). While this distinction is increasingly contested, it will be maintained here. It is, in any case, an open question as to just where the



boundary between the grammar and the lexicon lies and, more significantly for what follows, what kinds of 'basic irregularities' are possible; that is to say what kinds of idiosyncrasies can be expected to occur in the lexical items of a language.

Bloomfield's characterization raises the question as to the kinds of basic irregularities which might be found in the lexicon of a language. They appear to be of two types. Some irregularities are exceptional in cases where the grammar provides options but only one is taken in a particular lexical item. That does constitute an idiosyncrasy by way of arbitrary restriction but the rules of the grammar are not breached. Obligatory truncation is such a case as in (5) and (6). Phonetic truncation is an option the grammar provides but in (5) and (6) the MWEs are always truncated. Restricted collocations do not break any rules of the grammar. They provide arbitrary paradigmatic restrictions on linguistic choices when the grammar allows for a larger set of choices to be made. In such cases the grammar is permissive but the lexicon is restrictive. A second class of irregularities is the result of breaches of the grammar where the constraints imposed by the grammar do not provide alternatives. Here the grammar is restrictive but the lexicon is more permissive. Some borrowed words, for example, may breach the phonological constraints of a language. English phonotactics do not allow the onset sequence /ʃn/. But the dog breed of *schнауzer* is lexicalized in English with this initial cluster. Let us call exceptions of the first kind where they are manifest Type 1 idiosyncrasies and exceptions of the second kind Type 2 idiosyncrasies.

I now propose a hypothesis to link the grammar to its exceptions. The following hypothesis, the LAW OF EXCEPTIONS, is proposed as the strongest compatible with the distinction between the grammar and the lexicon.

LAW OF EXCEPTIONS: All formal properties of the grammar of a language are subject to exceptions manifested in idiosyncrasies in the lexical items of that language.

The Law of Exceptions thus predicts that the lexicon of a language will contain lexical items which break every rule in the grammar of that language. This is in line with the view of Di Sciullo & Williams (1987) with the metaphor that the lexicon is like a prison in that its inmates have all broken one or other law (although Di Sciullo and Williams do not suggest all laws are broken by at least someone).

Note that it cannot be assumed that the Law of Exceptions is *prima facie* true. It might well be that there are areas of the grammar of a language and perhaps all languages where there are no exceptions, i.e. that there are laws that are never

broken. Tensed verb second placement in main clauses of many Germanic languages is absolute as is suggested in the later discussion around examples (48)–(52). The verb second constraint in Dutch and German may be such an instance. The Law of Exceptions is therefore testable against all lexical items in the lexicon of a language.

This chapter focuses on the structural properties of a subset of lexical items, MWEs that have syntactic structure. Such lexical items may vary in many ways, so an account of the ways in which these properties can vary in general is useful, not least as a checklist for languages whose phraseology has not been documented. In the case of the languages of exemplification it will be shown that in every syntactic domain covered in this chapter, where there are regularities of both Type 1 and Type 2 in the grammar, there are exceptions in the lexicon.

Since the Law of Exceptions provides a relationship between the grammar of a language and its lexicon it is important initially to determine where exceptions cannot in principle be found. The prediction is that this will only be the case where a grammar has no regularities. For example, in the domain of morphology it is often considered that Chinese languages have no derivational morphology.¹ If that is the case, then the idiosyncratic properties manifested in the derivational morphology of derived words in other languages which do have derivational morphology are not in evidence in Chinese languages and so, obviously, are not available for analysis as to their idiosyncrasies. In the domain of syntax, since the syntax of a language determines what kinds of syntactic idiosyncrasies are possible, if the syntax of a language has antipassive voice, then there may be MWEs which exist only in the antipassive form or not in the antipassive form even when it is plausible that they should.² But, since only ergative languages have an antipassive voice, this syntactic property places a limit on the kinds of idiosyncrasies which can be expected in the lexical entries of lexical items in a particular language.

Since the Law of Exceptions applies to the current synchronic grammar, all MWEs which were historically unexceptional but are exceptions to the synchronic grammar fall under the Law of Exceptions. The reason for this is that native speakers of a language may be presumed to have internalized only the synchronic grammar of their native language (historical linguists excepted).

¹But see Starosta et al. (1997) for a contrary view.

²It is likely that, at least for some grammatical rules, there is more than one way in which they may be violated. Take for example, the English passive. It may be that an MWE is only possible with a *get* auxiliary and not a *be* auxiliary or that the agent which is in an oblique position in a passive MWE cannot be deleted. The Law of Exceptions, therefore, needs to note in how many ways a grammatical rule might be breached. It is an open question whether all possible breaches have associated MWEs.

Turning now to MWEs, if we suppose following Sag et al. (2002) that MWEs are lexical units of more than one word, then in this chapter the analysis of the properties of MWEs will be restricted to those of the subset of phrasal vocabulary, i.e. MWEs having syntactic structure. The Law of Exceptions predicts that the lexicon of a human language will always contain an inventory of MWEs with grammatical structure since all languages have syntax. Such lexical items are in the mental lexicon because they have one or more idiosyncrasies, i.e. properties which cannot be predicted by the grammar of the language. That is why they are stored and retrieved rather than computed (Bresnan 1981).³ Such lexical units are elsewhere termed, amongst other things, ‘phrasemes’ (Mel’čuk 2012) or are a subclass of ‘morpheme equivalent units’ (Wray 2008).

It follows that the structural properties of compound words will not be examined. This is because opportunities for structural variation in compounds are relatively slight (Selkirk 1982) while in the subset of MWEs with syntactic structure, opportunities for idiosyncratic variation of structural properties are considerable since the syntaxes of natural languages are complex and thus offer many opportunities for syntactic idiosyncrasy.

MWEs have two kinds of properties: digital properties such as having an obligatory plural in some instances, and gradable (analogue) properties such as their degree of semantic compositionality. Analogue properties can be of two kinds: the MWE has a particular property to a greater or lesser degree or the MWE has the property some of the times it is uttered but not at other times.

Viewed diachronically, MWEs may exhibit idiosyncratic properties that were not idiosyncratic at some time in the past. Some of the many English MWEs which are originally quotations from Shakespeare and the King James bible translation, often termed ‘winged words’ in continental phraseological manuals (Gläser 1986) have such idiosyncrasies. They are not alone as can be seen by (1).

- (1) *will he nill he*
 - a. originally: *will he ne will he*
 will he not will he
 - b. now truncated further to: *willy-nilly*⁴
 ‘regardless of what one might wish’

³That is not to say that they are unanalysable, as hybrid theories of speech production such as those of Cutting & Bock (1997), Titone & Connine (1999) and Sprenger et al. (2006) propose.

⁴Such archaisms have been noted in the inventory of MWEs for sources as various as Homeric epic (Lord 1960) and livestock auctions (Kuiper & Haggio 1984).

What general sources are there for the idiosyncratic properties of MWEs? The idiosyncratic properties of MWEs have three sources:

1. properties they have by virtue of being lexical items;
2. properties they have by virtue of being structurally complex;
3. properties they have by virtue of being phrases.⁵

The subset of MWEs which I will examine, as I have indicated above, may be defined on the basis of their structural properties, namely that they have syntactic structure. They may have other properties which may cross-classify. Sag et al. (2002) see idiomaticity as definitional for MWE. I do not regard semantic non-compositionality as a definitional criterion for MWE since it is shared with derived words (Jackendoff 2002) and, although many MWEs are idioms, many are not (Mel'čuk 2012). In example (2),

(2) *infidelity*

has a narrowed sense of 'marital infidelity'.

Having associated conditions of use is also a cross-classifying property since mono-morphemic words may also have associated conditions of use as in example (3).

(3) *Thanks!*

The property of being a restricted collocation is common to compounds and idioms. In a compound, since both constituents are lexicalized they must be restricted against one another.

That being the case an MWE can be a restricted collocation, semantically non-compositional, and have associated conditions of use as in example (4).

(4) *I declare the meeting open.*

Example (4) is an MWE. It is a restricted collocation. *Open* has a somewhat specialized sense and the whole expression is a formula used by the chair of a meeting to begin the formal proceedings of a meeting.

A classification of all lexical items on the basis of structural properties (which do not cross-classify) can be given as in Figure 1.⁶

⁵Here phrases are to be understood to include clauses and sentences, i.e. a sequence of words having syntactic structure.

⁶This is also the approach used by Fiedler (2007).

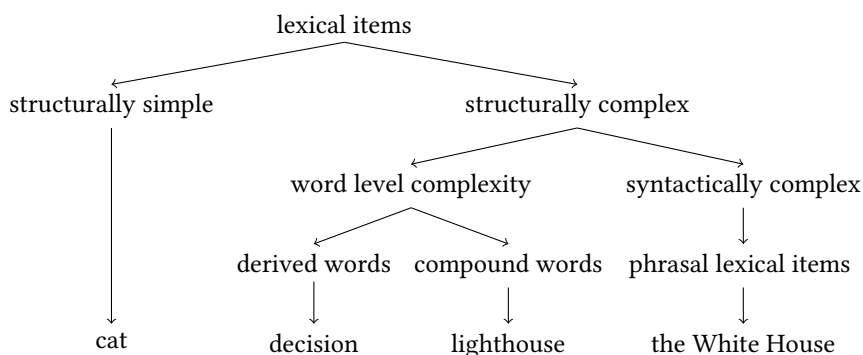


Figure 1: Structurally based classes of lexical items.

2 Idiosyncratic properties of MWEs

Many of the properties described below are described and exemplified for German by Burger (2010) and in Jaki (2014).

2.1 Idiosyncratic properties of MWEs which they have by virtue of being lexical items

Such properties are shared with structurally simple lexical items.

All MWEs may have phonological idiosyncrasies. This is probably a digital property. For example, MWEs may be lexicalized with idiosyncratic phonetic realizations, e.g. obligatory truncation as in (5). As noted above, this is a Type 1 idiosyncrasy.

- (5) *She'll be right.*
 #*She will be right.*

An Australian MWE indicating that there is nothing to be concerned about, and

- (6) *Good day.*

an Australian English greeting being conventionally realized as [gɪdeɪ].

MWEs may also have idiosyncratic intonation contours, e.g. livestock auction formulæ (Kuiper & Haggio 1984), market cries and classroom greetings by elementary school children to their teacher in Australia and New Zealand which go at half normal articulation speed and have a distinctive tune on the formula as in (7).

- (7) *Good morning, Miss/Mrs/Mr X.*



Good morning, Mister Jones

Figure 2: Primary school greeting formula tune.

- (8) Māori
Tihei mauri ora
sneeze spirit life
'the sneeze of life'

(8) is used by a speaker when taking the floor during Māori oratory.

The formula has an intonationally raised and prosodically drawn out syllable on *hei* and a quicker than normal downward intonation contour on the remaining syllables of the phrase. This is also a Type 1 idiosyncrasy since such intonation contours are possible within the grammar.

Any lexical item may have conventional conditions of use such as (9)–(11). This is a digital Type 1 property since the grammar has nothing to say about the usage conditions of lexical items. Such lexical items are often termed formulae or routine formulae (Coulmas 1979).

- (9) *Sorry.*

a single word apology,

- (10) *Bullshit!*

a compound word exclamation of disbelief,

- (11) *If it please Your Honour.*

an MWE used by legal counsel seeking approval for a course of action of a presiding judge in a court

An example from Māori is the formula in (12).

- (12) *kapiti hono, tātai hono*
join connect recite connect

This is a bridging formula used to transition from the acknowledgement of the dead to greetings to the living in formal speechmaking.

While I have designated this property as digital, i.e. an MWE either does or does not have specific conditions of use, the specific conditions of use are themselves complex and not necessarily digital (Biber 1994). The contexts of use will also range from the general to the specific.

2.2 Idiosyncratic properties MWEs may have by virtue of their being structurally complex

Such properties are shared by structurally complex words.

It is possible for a MWE to exhibit morphological and/or morphosyntactic idiosyncrasy. The extent to which this is possible depends on the inflectional and derivational morphology of a language.⁷ Languages with extensive inflectional morphology such as Turkish would be expected to exhibit inflectional idiosyncrasy in their MWEs. Chinese languages in the absence of inflectional morphology cannot. For example, in English an MWE may have an obligatory singular when a plural is semantically plausible as in (13) or an obligatory plural as in (14) when a singular is plausible. These are Type 1 idiosyncrasies.

- (13) *give someone a hand*
'assist someone'
#*give someone a pair of hands*
- (14) *as scarce as hens' teeth*
'very scarce'
#*as scarce as a hen's teeth*

MWEs may also have idiosyncratic derivational morphology. In the MWE in (15),

- (15) *the use of undue force*
or: *to use undue force*

⁷Note that it is rare for left hand constituents of English compounds to have inflections even when this is warranted semantically as in *head count* 'the counting of heads'. This restriction on the appearance of inflections appears to be a requirement of the word formation rules of English and thus not an idiosyncratic property of individual compounds.

this formula is used of police arrests in particular. The morphological idiosyncrasy is that there is no equivalent with *due force* (although there are no doubt situations where due force is applied).

In Māori, the expression

- (16) *mā-na (noa ake) te kore e V (o NP)*
 PREP-3SG (just up) DET NEG TAM V (PREP NP)
 Lit. (His/they, ...) not V-ing is (just) his fault.
 ‘NP is certain to V’

occurs in the example sentence in (17).

- (17) *Mā-na noa ake te kore e pai o tā tātou rārangi*
 PREP-3.SG just up DET NEG TAM good of SG.A L.PL.INC line
 ‘Just because of him our lineout was no good.’⁸

This is from a piece of talk about rugby football in which lineouts are a set move.⁹

This MWE has the restriction that it always contains *māna*, with 3rd person singular agreement, whatever the number of the subject NP. This is a case where the VP has a frozen morphology and agreement does not operate across the boundary between the open slot of the subject and the number inflection of the verb. Since the general rules of agreement do not operate in this case, this is a Type 2 idiosyncrasy.

An MWE may also retain as an idiosyncratic property an inflection which is no longer available in the current language. The Dutch MWE in (18)

- (18) *des duivels*
 the.GEN devil.GEN
 ‘to be very angry’

exhibits a genitive inflection on the article which is no longer current. This is a Type 2 idiosyncrasy since (18) exhibits a breach of the current rules of inflection.

It is also possible that an individual word may have different morphosyntactic properties in an MWE than it does elsewhere. An anonymous reviewer has given the following example from French. ‘To cite from Grevisse (*Bon Usage*: 198) “*Orge est féminin, sauf dans les deux expressions orge mondé, orge perlé.*”

⁸Detailed glosses: *Māna* ‘for him/her’; *noa ake* ‘just/merely’; *tā tātou* = ‘ours (first person inclusive)’, i.e., ‘belonging to all of us’.

⁹Source: http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=10587508.

So a word may be feminine, except in certain fixed expressions.’ Here presumably agreement with the gender of the noun would be un-idiosyncratic. This is a Type 1 idiosyncrasy since the grammar of French makes two genders available and the assignment of gender to individual words is (in part) arbitrary.

All structurally complex lexical items can have bound forms as constituents, derived words such as *agog*, compound words such as *wardrobe* (Richter et al. 2010). In MWEs the following examples show the presence of bound words:¹⁰

- (19) Māori
māngere hōnia
lazy very
‘very lazy’
- (20) *be on tenterhooks*
‘to be in a state of agitation about a future event’
- (21) *take umbrage at*
‘take offence at’
- (22) *kith and kin*
‘relatives’

Being a bound word is a Type 1 idiosyncrasy since no rule of the grammar is breached by the fact that a word is bound within an MWE.

While this chapter will not specifically deal with the semantic idiosyncrasy of MWEs, it is useful to offer a few remarks on that property since it is shared with structurally complex lexical items such as derived and compound words. Semantic idiosyncrasy appears to be an analogue property.¹¹ Semantic idiosyncrasy can come about in a number of ways. As Jackendoff (1975) points out, many English derivational affixes are polysemous. In particular words, however, only one of the senses associated with the affix is part of the compositional reading of the word as a whole. Such selective compositionality occurs where not all the cross-product senses of affixes or words are part of the sense of the whole expression. In an MWE such as (23),

- (23) *BE in stock*
‘be part of the current inventory of a shop or warehouse’

¹⁰In (19), *hōnia* is a bound form occurring only as a modifier of *māngere* ‘lazy’.

¹¹See Burger (2010) for a useful introductory discussion.

the word *stock* does not have the sense of ‘liquid used for soups and sauces’ but ‘inventory’.¹² This is a Type 1 idiosyncrasy since such a reading is a possible reading but not the only possible one.

Non-compositionality can occur where the sense of a word in a lexical item does not occur when the word is used independently. This is not a matter of a breach of a grammatical or semantic rule. It would therefore be a Type 1 idiosyncrasy. It is manifest in a MWE such as (24).

- (24) *without let or hindrance*
 ‘without any obstruction or interruption’

The noun *let* has a now defunct syntactic category and sense.¹³

A second kind of non-compositionality occurs when the rules of the semantics of the language are breached as in conventional figurative expressions such as (25). These are Type 2 idiosyncrasies. In such cases structurally complex lexical items also share the potential property of having analysable semantic representations. In (25), the phrase is figurative with the gloss ‘accepting a difficult challenge or situation’, *grasp* being ‘accept’ and the *nettle* being ‘the difficult challenge or situation’.

- (25) *grasp the nettle*

3 Idiosyncratic structural properties a MWE may have by virtue of being a phrase

In this section, I focus on those areas of potential structural idiosyncrasy which MWEs have but which they do not share with other structurally complex lexical items such as derived and compound words. Each relevant area will show that the Law of Exceptions appears to be corroborated.

All MWEs are associated with a phrase structural configuration. This is shown for one MWE in example (26), where the final NP is a slot (an open argument position).

- (26) [VP[V *make*][NP[DET *the*][N *most*][PP[P *of*][NP]]]
 ‘maximize the potential offered by ...’

¹²This selective compositionality may be the consequence of polysemy or homonymy. It can be difficult to separate these in particular instances.

¹³It also has this sense in the term *let* in tennis, namely an obstruction.

Having grammatical structure is a digital property. That is not to say that such structures are always permissible in the current synchronic grammar of the language as in (27).

- (27) *be that as it may*
'whatever the actual case may be'

Example (27) is in the subjunctive mood, a mood that no longer exists in the current synchronic grammar of English.¹⁴ This is a Type 2 idiosyncrasy.

In (28), the syntax is calqued from a Chinese four character idiom (Kuiper & Tan 1989).

- (28) *long time no see*
'I haven't seen you for some time'

It can therefore be concluded that there are exceptions to the phrase structural regularities of the synchronic grammar. This is also a Type 2 idiosyncrasy.

The phrase structure of MWEs may be further constrained by general constraints that are lexicon internal in that not all the possible phrase structural configurations the grammar allows are to be found in MWEs. For example, grammars allow for recursive rules in their syntax. There is evidence of a degree of recursion in MWE as in (29).

- (29) *a sight for sore eyes*
'a welcome appearance'

In (29), there are two NPs one is the top NP while the other is embedded in a PP. However recursion is limited because MWEs are stored in a finite brain and so cannot be of indefinite length. How limited recursion in MWEs may be is an open question.¹⁵ The Law of Exceptions is, however, corroborated as regards recursion in the grammar since this is a Type 2 idiosyncrasy albeit a general one since it is not just the property of a single lexical item.

O'Grady (1998) proposes that the citation form of idioms in the mental lexicon is in the form of lexical selection by heads of heads within their syntactic domains thus forming chains of heads. Some of these requirements are interesting in that, while phrases must have heads, it is not a necessary property of MWEs that the head position should dominate a lexical item or, in the case of functional projections, a specific functional head. This constraint may itself have exceptions,

¹⁴This is essentially a morphosyntactic property included here as a structural property.

¹⁵Hoeksema (2010), and Richter & Sailer (2009) discuss MWEs of clause length.

as suggested by an example from Māori as in (30), a formula expressing sympathy for someone's problem.

- (30) *i wāna nei (hoki)*
 at/by/from his/her NEAR-SPEAKER EMPHATIC-PARTICLE

While *i* is the head of the PP, O'Grady's theory predicts that there must be a lexical head for *i* to select within the immediate domain of the PP, i.e. the head of the NP complement of *i*. *Hoki* is not a lexical head and is optional (therefore cannot be a lexical head). *Wāna* 'his/her' can be replaced by an appropriate possessive determiner, e.g. *wāu* 'your', *wā rātou* 'their'. There are however restrictions on the choice of possessives. The possessive pronoun always starts with *w*, a possible initial phoneme for these possessives otherwise, but many speakers use it only here. Normally the possessive determiners have *t-* for singular possessum, \emptyset - for plural, thus eg *tāna* ~ *āna*. Some speakers allow *w-*, thus *wāna* for plural. However all speakers use only the *w*-forms in this MWE showing an idiosyncrasy typically associated with MWEs. Furthermore the possessive (taking possessive to be determiners) has no NP complement. The determiner position is also always a possessive, i.e., *i wōna nei* can not occur in this MWE. *I wōna nei* can certainly appear elsewhere given the right syntactic etc environment as in (31); just never in this MWE.

- (31) *He nui ake ōku waka i wōna nei.*
 A big upwards my cars than his here'
 'My cars are bigger than his.'

Nei is obligatory. While in the morphosyntax of Māori there are three locative particles, one never finds either of the other locative particles: *nā*, 'near hearer' or *rā* 'over there' in this MWE. So there are no lexical heads within the domain of the head position of this MWE and the non-heads are idiosyncratic in various ways. Thus, unless one uses an analysis allowing functional heads to serve in O'Grady's head chain proposal in which *wāna* etc. are Determiners and thus functional heads, this MWE has no lexical head within the immediate domain of the head of the MWE.¹⁶ This case therefore suggests that a strong form of O'Grady's proposal is falsified and the Law of Exceptions is corroborated.

¹⁶Ray Harlow (personal communication) provided this example and analysis. A case might be made for functional heads as well as lexical heads being predicted to be lexicalized in the case of possessives where the possessive marker could be regarded as a functional head of DP and where the NP within the possessive phrase is a slot.

The syntax of a language may require certain obligatory constituents, e.g. complements of transitive verbs or possessive NPs. In an MWE these may not be filled with lexicalized material as in (32) and (33). Where this is the case the idiosyncrasy is lexical and is not a breach of the rules of the syntax. It is an idiosyncrasy because, while *take* is transitive, it is an idiosyncratic feature of the MWE that the NP object of *take* in *take NP to task* is not lexically filled as it is in *take notice of*. Thus these are Type 1 idiosyncrasies.

(32) *take NP to task*
'hold someone responsible'

(33) *get NP's goat*.
'annoy someone'

Such slots may be semantically restricted in idiosyncratic ways as in (34).

(34) *drop in on NP[+human]*
'visit someone unannounced'

Some MWEs have an optional but lexicalized constituent as in (35). Again this does not involve a syntactic irregularity since the rules of the syntax allow both configurations. In other words, while *drop* can have both human and non-human objects, *drop in on* can only have a human object. So these are also Type 1 idiosyncrasies.

(35) (*keep poss-NP fingers crossed*)
'hope for a good outcome'

Such optional constituents are either truncations as in (34) or they can be internal to the MWE as in (36).

(36) *take (careful) note of NP*

In (36), *careful* is a highly preferred modifier and thus we can suppose that it is lexicalized but optional. Given that modifiers are permissible in general, this is a Type 1 idiosyncrasy.

The distinction between slots and optional constituents is that slots are lexically unspecified except for their syntactic category and are obligatory, optional constituents are lexically specified and optional while modifiable MWEs are optionally able to take any appropriate modifier.

Conversely, in some MWEs, internal modification having scope over an internal constituent is not permitted, for example in the case of (37).

- (37) *cut no ice*
 ‘have no impact’

This cannot be modified, as shown in (38),

- (38) #*cut no melting ice*

when the grammar would otherwise permit it.¹⁷ This suggests that modifiability properties can sometimes be absolute such as cases where modifiability is impossible whereas for other MWEs it may be a preferred option with some highly favoured and other disfavoured modifiers.

The presence of a lexicalized optional modifier in (36) is idiosyncratic but it is not a grammatical irregularity. It is therefore a Type 1 idiosyncrasy. The restriction against modification in (37) could, by contrast, be regarded as a grammatical irregularity, i.e. a Type 2 idiosyncrasy. That such cases should exist is a prediction of the Law of Exceptions.

Where the syntax of a language allows a variety of related constructions for a similar argument structure, an MWE may only permit one or fewer than a full set of variations, e.g. double object constructions as in (39)–(42), passives as in (43) and (44). This is a Type 1 idiosyncrasy.

- (39) *give NP the sack*
 ‘terminate NP’s employment’
- (40) #*give the sack to NP*
- (41) #*pay something attention*
- (42) *pay attention to something*
 ‘note or concentrate on something’
- (43) *NP cut off his/her nose to spite his/her face*
 ‘act in a way that is detrimental to oneself out of pique’
- (44) #*John’s nose was cut off to spite his face.*

¹⁷The semantics of modifier constituents within MWEs is complex (Nicolas 1995). Nicolas suggests that a modifier placed internally to a MWE can have scope over the meaning of the whole expression. In (38) the internal modifier *melting* cannot be parsed as modifying *ice*. However there are other cases Nicolas regards essentially as adverbial in having scope over the metaphorical expression as a whole so that *cut no real ice* is parsed as ‘really cut no ice’ and *cut no empirical ice* is parsed as ‘cut no ice empirically’. So the modification, while placed internally, is not semantically a modification of the constituent the modifier is predicted to modify in a compositional syntax. These cases are thus idiosyncratic. In that sense the placement of the modifier is structurally idiosyncratic.

For each set of syntactic alternates of this kind there may be MWEs which are idiosyncratic in allowing only one of the two possibilities where the grammar would predict that both might occur. For example, a double object construction may be lexicalized in either one or other form as in (39)–(42), or both as in (45) and (46).

- (45) *give credit to NP (for)*
'give positive acknowledgement to someone for something'
- (46) *give NP credit (for)*

Such distributions may well be matters of degree.¹⁸

In an MWE the antecedent of a pronominal or reflexive can be more restricted than the syntax of a language requires. Such cases can be seen as slots which are restricted to pronominals with additional slot restrictions as regards the antecedent of the pronominal. This is a Type 1 idiosyncrasy.

In (47), the antecedents of the possessives must be the agent arguments of *dig* as in (47a) and (47b).

- (47) *dig one's heels in*
'resist'
- a. *Jane dug her heels in.*
- b. # *Jane dug Fred's heels in.*

An MWE can have argument structure which is different from that of its head verb as in (48). This again a Type 1 idiosyncrasy since the grammar allows for predicates to have various argument structures.

- (48) *raining cats and dogs*
'raining heavily'

Rain is a zero place predicate but in (48) it is apparently a one place predicate.¹⁹

MWEs merge into the constructions of Construction Grammar. They are of two kinds: lexically motivated constructions, e.g. the *let alone* construction (Fillmore et al. 1988), and syntactically motivated constructions, e.g. irreversible binomials (Malkiel 1959). It is an open question whether the latter belong in the

¹⁸Fraser (1970) hypothesizes that there is a hierarchy of frozenness in construction types while Nunberg et al. (1994) propose that the degree of syntactic flexibility is related to the degree of compositionality of the MWE.

¹⁹This could be seen as a case of a lexicalized internal accusative such as one gets with *It snowed a blizzard.*

- (52) *Ik dacht dat ik het gewoon uit mijn duim zuigen kon.*
I thought that I it just out my thumb suck could
'I though I could just make it up.'
- (53) *Ik zoog het gewoon uit mijn duim.*
I sucked it just out my thumb
'I just made it up.'

What is the order in German and Dutch of the verb plus its complement for such MWEs in the mental lexicon? Is there an order at all or are there only dependencies? This is not a question of flexibility under movement. Verb second placement in main clauses is obligatory. Are there Type 2 idiosyncratic manifestations of these regularities in the MWEs of German and Dutch? The Law of Exceptions predicts that there should be cases in Dutch and German of MWEs which have main clauses where the tensed verb is not in second position.

Beyond the Law of Exceptions lies a further question as to the preponderance in the lexicon of a language of particular classes of exceptions. It is possible that different languages make different selective use of the parameters of variation noted above, e.g. some languages might have more bound words than others (Dobrovol'skij 1988).

4 Conclusion

The foregoing provides an outline of a set of structural properties of a grammar which have the potential to have exceptions and thus give rise to idiosyncratic structural properties of MWEs. It has been proposed that such exceptions are of two kinds: those which are 'basic irregularities', i.e. which are in breach of the rules of the grammar and which I have termed Type 2 idiosyncrasies, and idiosyncrasies which are the result of arbitrary restrictions on lexical items where the grammar makes no stipulation about such restrictions. Where such idiosyncrasies appear in lexical items, these have been termed Type 1 idiosyncrasies. By classifying idiosyncrasies on the basis I have, it also seems that Type 1 idiosyncrasies may be more common and diverse than Type 2 idiosyncrasies and that the lexicon is not as full of seriously lawless inmates as the prison metaphor in Di Sciullo & Williams (1987) suggests. Perhaps there is a maximum security wing for Type 2 inmates and a less secure set of cell blocks for Type 1 inmates whose deviance is by way of arbitrary restriction.

Acknowledgements

My considerable thanks to the editors of this chapter and its anonymous reviewers for numerous corrections and suggestions. I am grateful to Ray Harlow for the Māori examples. Note there has been no systematic study of the phraseology of Māori.

References

- Bargmann, Sascha & Manfred Sailer. 2018. The syntactic flexibility of semantically non-decomposable idioms. In Manfred Sailer & Stella Markantonatou (eds.), *Multiword expressions: Insights from a multi-lingual perspective*, 1–29. Berlin: Language Science Press. DOI:10.5281/zenodo.1182587
- Barkema, Henk. 1996. Idiomaticity and terminology. A multi-dimensional descriptive model. *Studia Linguistica* 50(2). 125–160.
- Biber, Douglas. 1994. An analytic framework for register studies. In Douglas Biber & Edward Finegan (eds.), *Sociolinguistic perspectives on register*, 31–56. New York: Oxford University Press.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart & Winston.
- Bresnan, Joan. 1981. A realistic transformational grammar. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 1–59. Cambridge, MA: MIT Press.
- Burger, Harald. 2010. *Phraseologie: Eine Einführung am Beispiel des Deutschen*. 4th edn. Berlin: Eric Schmidt Verlag.
- Coulmas, Florian. 1979. On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics* 3. 239–266.
- Cutting, J. Cooper & Kathryn Bock. 1997. That's the way the cookie bounces: Syntactic and semantic components of experimentally controlled idiom blends. *Memory and Cognition* 25(1). 57–71.
- Di Sciullo, Anna-Maria & Edwin Williams. 1987. *On the definition of word*. Cambridge, MA: MIT Press.
- Dobrovolskij, Dmitrij. 1988. *Phraseologie als Objekt der Universallinguistik*. Leipzig: Verlag Enzyklopädie.
- Fiedler, Sabine. 2007. *English phraseology: A coursebook*. Tübingen: Gunther Narr.
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64(3). 501–538.

- Fraser, Bruce. 1970. Idioms within a transformational grammar. *Foundations of Language* 6. 22–42.
- Gläser, Rosemarie. 1986. *Phaseologie der Englischen Sprache*. Tübingen: Max Niemeyer Verlag.
- Hoeksema, Jack. 2010. De localiteit van idiomen. <http://www.let.rug.nl/~hoeksema/localiteit>, accessed 2018-4-19. Manuscript, University of Groningen.
- Jackendoff, Ray. 1975. Morphological and semantic regularities in the lexicon. *Language* 51. 639–671.
- Jackendoff, Ray. 2002. What's in the lexicon? In Sieb Nooteboom, Fred Weerman & Frank Wijnen (eds.), *Storage and computation in the language faculty*, 23–58. Dordrecht: Kluwer.
- Jaki, Sylvia. 2014. *Phraseological substitutions in newspaper headlines: "More than meats the eye"*. Amsterdam/Philadelphia: John Benjamins.
- Kuiper, Koenraad & Douglas C. Haggio. 1984. Livestock auctions, oral poetry and ordinary language. *Language and Society* 13. 205–234.
- Kuiper, Koenraad & Daphne Gek Lin Tan. 1989. Cultural congruence and conflict in the acquisition of formulae in a second language. In Ofelia Garcia & Ricardo Otheguy (eds.), *English across cultures. Cultures across English*, 281–304. Berlin: Mouton de Gruyter.
- Lord, Albert Bates. 1960. *The singer of tales*. Cambridge, MA: Harvard University Press.
- Malkiel, Yakov. 1959. Studies in irreversible binomials. *Lingua* 8. 113–160.
- Mel'čuk, Igor. 2012. Phraseology in the language, in the dictionary, and in the computer. In Koenraad Kuiper (ed.), *Yearbook of Phraseology*, 31–56. Berlin: de Gruyter.
- Nicolas, Tim. 1995. Semantics of idiom modification. In Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schroeder (eds.), *Idioms: Structural and psychological perspectives*, 233–252. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nunberg, Geoffrey, Ivan A. Sag & Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538.
- O'Grady, William. 1998. The syntax of idioms. *Natural Language and Linguistic Theory* 16. 279–312.
- Richter, Frank & Manfred Sailer. 2009. Phraseological clauses in constructional HPSG. In Stefan Müller (ed.), *Proceedings of the 16th international conference on Head-Driven Phrase Structure Grammar, Göttingen*, 297–317. Stanford, CA:

- CSLI Publications. <http://csli-publications.stanford.edu/HPSG/2009/richter-sailer.pdf>, accessed 2018-4-19.
- Richter, Frank, Manfred Sailer & Beata Trawiński. 2010. The collection of distributionally idiosyncratic items: An interface between data and theory. In Stefaniya Ptashnyk, Erla Hallsteinsdóttir & Noah Bubenhofer (eds.), *Corpora, web and databases: Computer-based methods in modern phraseology and lexicography*, 247–261. Schneider: Hohengehren.
- Sag, Ivan A., 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 Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, 1–15.
- Schenk, André. 1995. The syntactic behavior of idioms. In Martin Everaert, Eric-Jan van der Linden, André Schenk & Ron Schreuder (eds.), *Idioms. Structural and psychological perspectives*, 253–271. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Selkirk, Elizabeth. 1982. *The syntax of words*. Cambridge, MA: MIT Press.
- Sprenger, Simone A., Willem J. M. Levelt & Gerard Kempen. 2006. Lexical access during the production of idiomatic phrases. *Journal of Memory and Language* 54. 161–184.
- Starosta, Stanley, Koenraad Kuiper, Siew Ai Ng & Zhi-qian Wu. 1997. On defining the Chinese compound word: Headedness in Chinese compounding and Chinese VR compounds. In Jerome P. Packard (ed.), *New approaches to Chinese word formation: Morphology, phonology and the lexicon in Modern and Ancient Chinese*, 347–370. Berlin: Mouton de Gruyter.
- Titone, Debora A. & Cynthia M. Connine. 1999. On the compositional and non-compositional nature of idiomatic expressions. *Journal of Pragmatics* 31. 1655–1674.
- Wray, Alison. 2008. *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.

