
Part 2: **Meaning (semantic relations)**

Steve Pepper

Hatcher-Bourque: Towards a reusable classification of semantic relations

Abstract: A key feature of binominal lexemes is the unstated (or underspecified) relation, $\mathfrak{R}$, that pertains between the two major constituents. The nature of $\mathfrak{R}$ – the kinds of relations – has been the topic of considerable research during recent decades. While early studies focused almost exclusively on English, the last few years have seen a spate of work on other languages. Unfortunately, this work has been uncoordinated and each researcher entering the field has tended to devise their own classification, making it difficult to compare results and advance our understanding of the phenomenon. This is a pity, because such an understanding has the potential to provide insights into the nature of concept combination and the associative character of human thought. The purpose of this chapter is to present a well-documented, systematic classification of semantic relations that operates at multiple levels of granularity and is suitable for reuse across languages. Hatcher-Bourque is based on revisions of two earlier classifications, those of Anna Granville Hatcher and Yves Bourque, which operate at different levels of granularity. These are integrated into a single, coherent system, with automatic mapping from one level to the other. The classification is applied to a set of 3,650 binominals from 106 languages, and an analysis is presented of the frequency and distribution of semantic relations at both a highly abstract level and a more granular level. The Hatcher-Bourque classification, and an accompanying, Excel-based tool, the *Bourquifier*, are offered to the research community in order to encourage collaboration, and researchers are invited to participate in the Hatcher-Bourque Cake Challenge.

1 Introduction

1.1 Background

The unstated (or underspecified) semantic relation, $\mathfrak{R}$, is a defining feature of binominals (see *Introduction*). Jackendoff (2016) provides a nice set of examples to show that the kind of semantic relation can be “hugely varied”, even across binominals that share a common head, such as *cake*.

Note: This chapter has been made Open Access *in memoriam* my parents Harry Pepper (1926–1996) and Edna Pepper (1932–2022).

<i>chocolate cake</i>	‘a cake made with chocolate in it’
<i>birthday cake</i>	‘a cake to be eaten as part of celebrating a birthday’
<i>coffee cake</i>	‘a cake made to be eaten along with coffee and the like’
<i>marble cake</i>	‘a cake that resembles marble’
<i>layer cake</i>	‘a cake formed in multiple layers’
<i>cupcake</i>	‘a little cake made in a cup’
<i>urinal cake</i>	‘a (nonedible) cake to be placed in a urinal’

The nature of $\mathfrak{N}$ has been a perennial topic of interest in the study of compounding that can be traced back to the Sanskrit grammarians. Modern treatment of the topic can be said to originate with Jespersen’s (1942) discussion in Volume 6 of his *Modern English Grammar on Historical Principles*.¹ The year 1960 saw the publication of three seminal works by Marchand, Lees and Hatcher that inspired further work in a number of different directions. In the years that followed there were important contributions by Adams (1973), Downing (1977), Levi (1978), Warren (1978), Ryder (1994), Jackendoff (2009; 2010; 2016) and Schäfer (2018), all of which focused on English [ENG]. More recently the topic of semantic relations has been explored in other languages, including French [FRA] (Arnaud 2003; 2016; Bourque 2014), Nizaa [SGI] (Pepper 2010), Danish [DAN] (Szubert 2012), Norwegian [NOR] (Eiesland 2016) and Spanish [SPA] (Toquero 2018). The matter has also received considerable attention in computational and corpus linguistics (e.g. Vanderwende 1994; Moldovan et al. 2004; Girju et al. 2005; Ó Séaghdha 2008; Tratz & Hovy 2010; Nakov 2013; Schäfer 2018) and was the focus of an NAA-CL-HLT Workshop on Semantic Evaluations task on “the interpretation of noun compounds using paraphrasing verbs and prepositions” (Butnariu et al. 2009).

1.2 Towards a reusable classification

The point of departure for the present chapter is three observations regarding this previous work. *The first observation* is that opinions differ as to whether the set of semantic relations found in binominals is finite or infinite. Jespersen (1942: 143) asserted that “the number of possible logical relations between the two elements [of a noun-noun compound] is endless” and Downing (1977: 810) concluded that “the semantic relations that hold between the members of [novel] compounds cannot be characterized in terms of a finite list of ‘appropriate compounding relationships’.” However, most researchers have had enough faith in the useful-

¹ But see also Grimm (1826), Mätzner (1860), Bergsten (1911) and Carr (1939).

ness of a finite list that they have taken the trouble to develop one. The position taken in the present research accords with that of Tratz & Hovy (2010: 679), who contend that “the vast majority of noun compounds fits within a relatively small set of categories.” Furthermore, it seems likely that, while the interpretation of novel compounds depends greatly on context, established compounds do so to a lesser degree and are more likely to exhibit a fixed set of basic relations.

The second observation is that among authors who have attempted to enumerate a list of relations, the number of relations varies considerably from four (in the case of Hatcher), to upwards of 40 or 50 (depending on whether or not subtypes are included). The position taken in the present paper is that the number of relations one identifies should be a function of the degree of granularity required by the investigation in question. It can therefore be anything the researcher desires, from one (as suggested by Bauer 1979) to unlimited (as opined by Jespersen). We further claim that any relation can be subdivided into more specific relations, if the need arises and – concomitantly – that any two arbitrary relations can be combined into a single, more general relation.

For some investigations, a small number of (high-level) relations will suffice; for others, a larger number of (low-level) relations is required. The advantage of a granular, low-level classification is that it is more concrete, and thus much easier to apply in practice; its disadvantage is that it results in a rather fragmentary picture from which it can be difficult to generalize. The advantage of a more abstract, high-level classification is that the generalizations are built into the scheme; its disadvantage is that the high level of abstraction makes it extremely hard to apply consistently.

This suggests that a classification scheme that operates at more than one level of granularity – with automatic mapping from lower to higher levels – may prove beneficial. Such a scheme, if based on sound principles, would enjoy both of the advantages outlined above, and suffer from neither of the disadvantages.

The third observation that is relevant here is that each researcher tends to construct their own scheme instead of reusing an existing one. That is the case in almost every one of the studies listed above, and one might legitimately ask why this should be so. Three possible reasons might be put forward. The first is simply that the material in question is notoriously slippery. Meaning only exists in our minds and is therefore hard to pin down. Getting inside someone else’s head is not easy, and it is made more difficult by the fact that many systems are rather poorly documented. The second reason is that judgements regarding the nature of a semantic relation are subjective and dependent on the level of granularity one aspires to: some might regard a system of 12 relations (such as Levi’s) as too vague, while others (Hatcher, no doubt) would find it too low-level (and too unsystematic). The third reason is that no system is perfect. It is easy to spot

inconsistencies and errors in others' work, and when we encounter such errors, there is a tendency to think that we can do a better job ourselves.

Whatever the reasons may be, the practice of discarding the work of others and starting from scratch does not seem conducive to the advancement of science. The position taken in this paper is that a better approach is to build on the work of earlier researchers, to reuse existing schemes, testing and refining them as necessary, and working incrementally towards the goal of a robust, flexible and easily reusable system that has been tested against different kinds of data from a large range of languages. The Hatcher-Bourque classification presented here is such a system. It is hereby offered to the research community as a basis for further collaborative work, together with an Excel-based tool for the computer-assisted analysis of semantic relations, the *Bourquifier* (Pepper 2021).

1.3 A note on terminology

Before proceeding, it is worth spending time to understand the structure of a semantic relation and the terminology to be used in this chapter.

The **relation** $\Re$ that pertains between the two major **constituents** of a binominal lexeme, such as *honey bee*, is by definition **binary**. It involves two **participants**, *honey* and *bee*, each of which plays a particular **role** in the relation. We can characterize the relation here as one of PRODUCTION: a honey bee is a bee that *produces* honey; the bee plays the role of producer and the honey plays the role of product.

It is important to distinguish between the role of a participant in a particular relation and its **type**, the class to which it belongs and that reflects its essential being. A bee is primarily an insect, not a producer, and honey is a kind of sweet fluid rather than just a product. Roles vary depending on the relation in question, whereas types are constant: the bee in *beehive* is playing a quite different role from the bee in *honey bee*, but it is still a bee.

All binary relations are **bidirectional**, in the sense that if A is related to B, then B is perforce related to A. In a **symmetric** relation, such as that of COORDINATION, B is related to A *in the same way* as A is related to B (if A is coordinate with B, then B is coordinate with A). In such relations, the role is the same for both participants. In an **asymmetric** relation, such as that of PRODUCTION, B is related to A *in a different way* from how A is related to B, and there are two distinct roles.

When a relation is asymmetric, it can take two forms depending on how the relation is profiled: in *honey bee*, the constituent denoting the producer (*bee*) is the semantic **head** and the constituent denoting the product (*honey*) is the **modifier**. By contrast, in *beeswax*, the constituent denoting the product is the head and the constituent denoting the producer is the modifier. Because it is asym-

metric, the PRODUCTION relation can be said to consist of two “**sub-relations**”, which we might label “producer of” and “produced by”. When both sub-relations are employed in binominal word-formation, the relation is said to be **reversible**, and the terms **basic** and **reversed** may be employed to distinguish between the two. Note that a relation may be asymmetric without necessarily being reversible.

In the following discussion, RELATIONS are shown in small caps and roles are underlined.

1.4 Structure of this chapter

This chapter is structured as follows: Following this introduction, §2 presents the low-level classification of 25 relations developed by Bourque (2014) that was chosen as the starting point for the present study; it also details the minor adjustments and extensions that were made to it, resulting in the Bourque29 component (29 refers to the number of relations) of the Hatcher-Bourque classification. §3 describes Hatcher’s (1960) high-level classification of four relations and how it was extended by the addition of one more relation in order to cover appositional as well as non-appositional binominals, resulting in the Hatcher5 component of the Hatcher-Bourque classification.

§4 describes the two-tiered Hatcher-Bourque system that results from the integration of Bourque29 with Hatcher5, and how this system relates to Aristotle’s three principles of remembering. §5 then presents the *Bourquifier* application and its use as a computer-assisted tool to expedite the analysis of semantic relations and ensure more consistent results.

§6 contains a statistical analysis, showing the frequency of various low- and high-level relations in a sample of 3,650 binominals from 106 languages, and §7 provides a conclusion and a challenge. Documentation for the complete Hatcher-Bourque classification is to be found in the appendix, in the form of detailed summaries of each relation and a one-page at-a-glance table.

2 Bourque’s low-level classification

2.1 Description of Bourque25

Out of the dozens of classification schemes to be found in the literature, the one selected for the present study is the one developed by Yves Bourque in his 2014 dissertation *Toward a typology of semantic transparency: The case of French com-*

pounds (Bourque 2014). This choice was dictated by a number of considerations, in particular the quality of Bourque’s documentation, which includes templates, linking material, examples and extensive discussion of overlaps between relations. A further reason was that the scheme avoids the Anglocentrism of many earlier studies, for example by providing examples in both English and French, employing descriptive labels (e.g. *PURPOSE* instead of Levi’s *FOR*), and using the terms ‘non-head’ (or ‘modifier’) and ‘head’ instead of the word order dependent ‘A’ and ‘B’ of Jespersen and Hatcher, or ‘N1’ and ‘N2’ of Levi and Jackendoff. In addition, a study involving nearly 4,000 binominals (Pepper 2020) shows that this classification operates at a level of granularity that is both *manageable*, in terms of the number of relations (25), and *precise*, in terms of expressing the nature of the various relations.

Bourque’s classification is furthermore based explicitly on a synthesis of 16 earlier classifications.² Whereas all but one of these are based on data from English, Bourque himself tested the system using a large database of French compounds, thus increasing the chance of cross-linguistic coverage.

From the 16 earlier classifications, Bourque synthesizes a set of “retained relations”, 15 in all, shown in Table 1. Of these 15, ten are considered to be reversible and are indicated by ^R.

Table 1: Bourque’s (2014:170) retained relations.

COORDINATION	COMPOSITION ^R	TIME ^R
HYPERNYMY ^R	SOURCE ^R	TOPIC
SIMILARITY	PART ^R	FUNCTION
PRODUCTION ^R	LOCATION ^R	PURPOSE
CAUSE ^R	POSSESSION ^R	USE ^R

Each relation is introduced by a summary table such as that exemplified for *PRODUCTION* in Figure 1. For reversible relations like *PRODUCTION*, the summary table consists of two rows, one for each of the (directed) ‘sub-relations’; these are labelled Basic and Reversed. Each row then contains a “structure template” in both English and French, examples (in the form of compounds, i.e. binominals of type **cmp** or **jxt**) from each language, and “linking material”. For non-reversible relations the second row is empty.

² Those of Jespersen (1942), Hatcher (1960), Adams (1973), Levi (1978), Downing (1977), Warren (1978), Shoben (1991), Vanderwende (1994), Lauer (1995), Rosario and Hearst (2001), Arnaud (2003), Moldovan & al (2004), Girju & al (2005), Girju & al (2009), Séaghdha (2008), Jackendoff (2010).

The structure template consists of a test frame with slots for the head (H) and modifier (M), respectively. Populating these slots with the constituents of a binominal results in a paraphrase of the relation that helps the analyst judge whether that relation is appropriate to the binominal in question. Thus we see that the paraphrase of *honey bee* as “a bee that makes honey” (the Basic form of PRODUCTION) provides a satisfactory reading, whereas that of the Reversed form, “a bee that honey makes”, does not. Conversely, *beeswax* is “a wax that bees make” and not “a wax that makes bees”.

PRODUCTION			
Relation Type	Structure Template	Examples	Linking Material
Basic	an H that makes M <i>un T qui fait M</i>	<i>honey bee</i> <i>appareil photo</i>	makes, produces <i>fait, produit</i>
Reversed	an H that M makes <i>un T que M fait</i>	<i>beeswax</i> <i>jazz manouche</i>	

Figure 1: Bourque’s template for PRODUCTION.

The linking material “is meant to draw parallels between the retained relation and those proposed elsewhere in the literature and may include such items as verbs (e.g. *have*, *cause*, *make*, etc.), prepositions (e.g. *for*, *from*, *of*, etc.), and even nouns (e.g. *kind*, *type*)” (Bourque 2014: 178).

In addition to the summary table, each relation is accompanied by a lengthy discussion that can run to several pages. This covers the precise nature of the relation, the ways in which it has been treated by earlier researchers, overlaps with other relations, and any other issues. The complete classification is summarized in Table 2.

Bourque’s system of 15 relations (of which 10 are reversible, for a total of 25 “sub-relations”) was sufficient to cater for the varied sample of nearly 4,000 binominals that will be described in §5. However, a few infelicities were discovered in the process, and when it came time to map the system to that of Hatcher, certain extensions were deemed necessary. The following section (§2.2) describes the non-substantive changes that were made to the original system, and §2.3 describes the substantive extensions that resulted in the Bourque29 component of the Hatcher-Bourque classification.

2.2 Non-substantive changes to Bourque

The non-substantive changes to Bourque’s classification involved renaming some relations, rewording some templates, and changing some examples. They are

presented in the following sections. (Refer to Table 2 for the original formulations and Table 7, in the Appendix, for the revised version.)

Table 2: The original Bourque25 classification.

Label	Type	Template	Linking material	Example
HYPERNYMY	Basic	an H of kind M	kind of,	<i>oak tree</i>
	Rev.	an H that M is a kind of	type of	<i>bear cub</i>
COORDINATION		a C is an H and an M	is also, is both / and	<i>boy king</i>
SIMILARITY		an H that is similar to M	similar to, like	<i>ant lion</i>
FUNCTION		an H that serves as M	functions, serves as	<i>buffer state</i>
POSSESSION	Basic	an H that possesses M	possess	<i>career girl</i>
	Rev.	an H that M possesses	(have / of)	<i>family estate</i>
PART	Basic	an H that is part of M	part of	<i>table leg</i>
	Rev.	an H that M is part of	(have / of)	<i>wheelchair</i>
LOCATION	Basic	an H located at/near/in M	at, near, in, etc.	<i>window seat</i>
	Rev.	an H that M is located at/ near/in		<i>bedroom</i>
COMPOSITION	Basic	an H made of M	composed/ made of	<i>sugar cube</i>
	Rev.	an H that M is made of		<i>sheet metal</i>
SOURCE	Basic	an H (made) from M	(made) from	<i>cane sugar</i>
	Rev.	an H that M is (made) from		<i>sugar cane</i>
CAUSE	Basic	an H that causes M	causes	<i>sunburn</i>
	Rev.	an H that M causes		<i>motion sickness</i>
PRODUCTION	Basic	an H that makes M	makes, produces	<i>honey bee</i>
	Rev.	an H that M makes		<i>beeswax</i>
TOPIC		an H about M	about	<i>history conference</i>
TIME	Basic	an H that occurs at/ during M	during, at, in, before, etc.	<i>summer job</i>
	Rev.	an H at/during which M occurs		<i>golf season</i>
USE	Basic	an H that uses M	use / with, by	<i>steamboat</i>
	Rev.	an H that M uses		<i>hand brake</i>
PURPOSE AND PROPER FUNCTION		an H intended for M	for	<i>animal doctor</i>

Changes to names of relations

Previous researchers have employed a variety of strategies for naming relations. Levi used a mixture of verbs (BE, HAVE, MAKE, CAUSE, USE) and prepositions (IN, FOR, FROM, ABOUT); Warren preferred to use role pairings (SOURCE-RESULT, WHOLE-PART, PART-WHOLE, SIZE-WHOLE, GOAL-OBJ, PLACE-OBJ, TIME-OBJ, ACTIVITY-ACTOR), but had recourse to other means for symmetric and non-reversible relations (COPULA, RESEMBLANCE, PURPOSE); Jackendoff's "basic functions" employ a verb-based naming system for the most part, but with the odd adjective, role or abbreviation thrown in (CLASSIFY, BE, BE AT/IN/ON, MADE FROM, CAUSE, MAKE, SERVES AS, HAVE, PROTECT (FROM), but also SIMILAR, KIND, PART, COMP).

Bourque's system is more consistent, but not entirely so. As Table 1 shows, all 15 relations are named by nouns. Of these, six are nominalizations of the verb or adjective typically used to express the relation, e.g. PRODUCTION < *produce* (the others are COORDINATION, COMPOSITION, POSSESSION, LOCATION and SIMILARITY). HYPERNYMY also denotes a relation, but one that is lexical rather than conceptual. PART, CAUSE, SOURCE, TIME and TOPIC, on the other hand, all denote one of the roles in the relation, while FUNCTION, USE and PURPOSE³ can denote either a role or a relation.

For the Hatcher-Bourque classification, it was considered desirable that names should denote relations rather than roles or linguistic means of expression, preferably using nominalizations of relevant verbs. Where this was not possible, a role-pair was preferred to a single role, but the latter was considered acceptable for symmetric and non-reversible relations. While it was not possible to achieve complete consistency, the following improvements were made:

- HYPERNYMY (lexical relation) > TAXONOMY (conceptual relation)
- PART (role) > PARTONOMY (relation)
- CAUSE (role) > CAUSATION (relation)
- TIME (role) > TEMPORALITY (relation)
- SOURCE (role) > SOURCE-RESULT (relation)
- USE (conversion) > USAGE (nominalization)

³ Bourque actually uses the name PURPOSE AND PROPER FUNCTION for this relation, but since his system already includes a relation called FUNCTION the name has been shortened. Considerations relating to having been designed to (or supposed to) perform a certain function (Millikan 1984: 17, cited in Jackendoff 2016: 23) are thus relegated to the description of the relation instead of its name.

The names TOPIC, FUNCTION and PURPOSE, on the other hand, were retained. Since these appear to be non-reversible, this is not a major issue, especially since the latter two can denote relations as well as roles.

Changes to templates

Changes made to Bourque's templates were motivated by the desire for greater transparency and/or consistency. The template for Reversed HYPERNYMY is actually incorrect, generating "a bear cub is a cub that bear is a kind of" for Bourque's example *bear cub*. A bear is a kind of animal, and not a kind of (bear) cub. One way to correct this error would be to adapt Jackendoff's "an N_2 that is a kind of N_1 " (i.e. "an H that is a kind of M"), but a simpler solution was to replace both HYPERNYMY templates with the more transparent "(an) M is a kind of H" and "(an) H is a kind of M". Thus, an oak is a kind of tree (whereas a tree is not a kind of oak), and a cub is a kind of bear (whereas a bear is not a kind of cub).

In addition, Bourque's templates for COMPOSITION, PRODUCTION and SOURCE (all of which employ the verb 'make') were modified to use the verbs 'compose' and 'produce' and the noun 'source', respectively, thereby tying the templates more closely to the name of the relation. For example, the template for Basic PRODUCTION (e.g. *honey bee*) was changed from "an H that makes M" to "an H that produces M", and that for Basic COMPOSITION (e.g. *sugar cube*) was changed from "an H made of M" to "an H composed of M".

Changes to examples

Most of the changes to Bourque's examples were motivated by pedagogical and, in a couple of cases, aesthetic considerations. Only one of his 25 examples is actually erroneous: the use of *sunburn* to exemplify Basic CAUSE, paraphrased as "an H that causes M". It is, of course, the sun (M) that causes the burn (H), not the other way round, so this example properly belongs under Reversed CAUSE, with the paraphrase "(a) burn that (a) sun causes". A better example for Basic CAUSE is *tear gas*: "(a) gas that causes (a) tear".

It is arguable that the example provided by Bourque for SIMILARITY is not incorrect, but it is certainly suboptimal. An *ant lion* (or antlion) is not a lion that is similar to an ant, it is a kind of insect, albeit not exactly an ant. The name appears to be a left-headed calque from Latin *formicaleo*, which means that the paraphrase "an ant that is similar to a lion" does in fact work. However, as a highly exceptional left-headed compound it is unsuitable in an English context for ped-

agogical reasons (it works fine as Fr. *fourmi-lion*, which may be how Bourque got to choose it as his example). It is replaced by *kidney bean* (a bean shaped like a kidney), an example taken from Hatcher (1960).

Conservation of space is the main consideration for choosing *history book* instead of *history conference* for TOPIC, and *sunburn* instead of *motion sickness* for Reversed CAUSE; real estate is at a premium not only on paper, but also in the *Bourquifier* (see §4.2).

Finally, a number of changes were motivated by the desire to use what David-Antoine Williams⁴ has dubbed “boathouse words” wherever possible, for pedagogical reasons. These are pairs of words that have the pleasing property of consisting of the same two constituents in reverse order, like Bourque’s examples for SOURCE, *cane sugar* and *sugar cane*. For COMPOSITION Bourque already has *sugar cube*, which is complemented nicely by *cube sugar*. In addition, *song bird* and *bird song* work well for PRODUCTION, as do *oil lamp* and *lamp oil* for USE, and *car motor* and *motor car* for PART. Finding a suitable boathouse pair for LOCATION is more difficult (the closest I have come is *house music* and *music hall*), and candidates are still being sought for POSSESSION, TIME, CAUSE and DIRECTION.

2.3 Substantive changes to Bourque’s classification

The more extensive changes to Bourque²⁵ involved the addition of codes for relations, the provision of names for roles, the enforcement of consistency when distinguishing between Basic and Reversed forms, and the addition of two new relations. These changes are described and justified in the following sections.

Addition of codes

In order to represent the data in a database and perform the quantitative study described in §5, unique identifiers were needed for each relation. Bourque will have encountered the same need, but he did not publish his codes, so new ones were created. These take the form of three- or four-letter mnemonic codes for Basic relations, and the same codes suffixed with -R for reversed relations, thus POSS for Basic POSSESSION, POSS-R for Reversed POSSESSION, etc. These codes are included in the documentation in the Appendix, in order to promote interoperability and to save other users the trouble of devising their own codes.

⁴ <https://thelifeofwords.uwaterloo.ca/boathouse-words/> (accessed 2021-02-10).

Addition of explicit roles

A more substantive change is the introduction of explicit names for the roles played by the participants in each relation, as described in §1.2. Thus, the PRODUCTION relation is supplied with the roles product and producer, the POSSESSION relation with the roles possessor and possessum, etc. Every asymmetric relation (reversible or not) involves two distinct roles, as here. The two symmetric relations, on the other hand, each involve a single role which is played by both participants in the relation. For COORDINATION that role is named coordinand, and for SIMILARITY it is named likeness.

These names serve as an aid in conceptualizing asymmetric relations, in understanding the difference between a Basic and its corresponding Reversed relation, and in describing and communicating about individual relations. If one adopts the convention of using the role played by the modifier to characterize a (directed) ‘sub-relation’, every one of the 29 sub-relations of the revised Bourque29 system can be referred to simply as the ‘X relation’. Thus, ‘possessor relation’ and ‘possessum relation’ can be used instead of the unwieldy terms Basic POSSESSION relation and Reversed POSSESSION relation for *family estate* and *career girl*, respectively. The reason why this works is because it proved possible to ensure that every role was unique – except for the use of entity as one of the roles in the TOPIC, FUNCTION and PURPOSE relations. Since these relations are non-reversible, there will seldom be a collision between multiple, homonymous *-entity relations.⁵

Alignment of Basic and Reversed relations

Asymmetric relations, as already noted, can take two forms, which work in opposite directions to one another (and incidentally are often paraphrased using active and passive sentences, respectively). As we have seen, Bourque labels these two forms Basic and Reversed, respectively, but he does not provide any rationale for choosing one form rather than the other to designate as Basic. It seems that the choice was essentially arbitrary. While this clearly did not matter to Bourque for the purpose of his investigation, such arbitrariness is unnecessary, leads to a less logically consistent result, and may prove confusing.

⁵ If at some point Reversed forms of TOPIC, FUNCTION and PURPOSE are found, the classification (including the relevant role names) will have to be revised.

Consider the relations PART, LOCATION and COMPOSITION in Table 2. All of these are in some sense specializations of a more general relation CONTAINMENT. If we now focus on the three Reversed examples of these relations (*wheelchair*, *bedroom* and *sheet metal*), we see that a part (*wheel*) is in some sense “contained in” its whole (*chair*), that a thing located (*bed*) is “contained in” its location (*room*), and that a material (*metal*) is “contained in” the object of which it is made (*sheet*). The wheel, the bed and the metal are thus all “containees” (in a very general sense), whereas the chair, the room and the sheet are all “containers”. However, while wheel and bed are denoted by modifiers, metal is denoted by the head constituent. If we now consider the POSSESSION relation (exemplified by *family estate*), in which the possessor somehow “contains” the thing possessed, we see that the containee *estate*, like *metal* (but unlike *wheel* and *chair*), is denoted by the head constituent.

This inconsistency can be removed by simply inverting Bourque’s POSSESSION and COMPOSITION relations such that the containees are denoted by the modifier instead of the head. Thus, the original pair of POSSESSION (sub-)relations (*before*) becomes the revised pair of (sub-)relations (*after*).

before:

POSSESSION	Basic	an H that possesses M	possess	<i>career girl</i>
	Rev.	an H that M possesses	(have / of)	<i>family estate</i>

after:

POSSESSION	Basic	an H that M possesses	possess	<i>family estate</i>
	Rev.	an H that possesses M	(have / of)	<i>career girl</i>

The same applies to COMPOSITION (and TAXONOMY) where Basic and Reversed are likewise inverted.

Similar considerations apply to Bourque’s SOURCE and USE, which are incompatible with CAUSE and PRODUCTION. Focusing again on the Reversed relations, we see that in the latter two relations, the participant that constitutes the point of origin (the motion in *motion sickness* and the bee in *beeswax*) is expressed by the modifier, whereas in the SOURCE and USE examples, the point of origin (the cane of *sugar cane* and the brake in *hand brake*) is expressed by the head. In the revised Bourque classification, the sub-relations of SOURCE and USE are therefore also inverted.

For consistency with other containment-related relations, forms such as *history book* are deemed to embody the Reversed form of the TOPIC relation rather than the (unattested) Basic form. Of the two paraphrases available for the CONTAINMENT relation, the Reversed form is clearly the most felicitous: cf. “a book that contains history” vs. the Basic form “*a book that is contained in (a) history”.

The attested forms of the other non-reversible relations, PURPOSE and FUNCTION, embody the Basic forms of the relation.

Addition of new relations

Although Bourque's set of 15 relations was sufficient to cater for the 3,650 binominals examined for the study on which this work is based, two more relations were added for the sake of completeness, and to facilitate the integration with Hatcher's system (described below). These were CONTAINMENT and DIRECTION.

The first of these was prompted by consideration of the Hawaiian binominal *pahu meli* [box honey] BEEHIVE. Is a beehive "a box that honey is part of" (Bourque's Reversed PART) or "a box that honey is located at/near/in" (his Reversed LOCATION)? Of course, location is involved, and one could also (at a pinch) say that honey is part of the beehive, but it would be more felicitous to say that the box *contains* honey. Now, although CONTAINMENT is not one of Bourque's relations, he does not ignore the matter. He discusses it in depth in the context of the overlap between PART and LOCATION, using the example of *toolbox*. His discussion is quoted here at some length in order to convey the detail of his discussions in general:

Another issue to consider is that some compounds might be analysed as either PART or LOCATION. This dual analysis is related to the fact that LOCATION may subsume PART: if something is a part of something else, then it is located at/on/in that thing (cf. Baron & Herslund 2001). One possible solution is to reserve location for only those compounds that actually involve a locative noun, as does Adams (1973). The problem, of course, is that one must treat combinations such as *toolbox* or *treehouse* using some other relation, as they do not, in the strictest sense, involve places. The key distinction that will be used here is one that views the PART relation as a reference to an integral component of the whole, without which it would either be incomplete, defective, or non-functional. Thus, a negation test may be used to determine whether the modifier denotes an essential part of the compound. The formulation in (105) below shows how such a test might apply to compounds in which the head denotes the whole (cf. 104 above):

- (105) a. a C without an M is still a C
b. un C sans M est toujours un C

A positive response to the above sentence would indicate that the modifying noun is not an essential component of the object denoted by the compound, but instead a distinguishing feature. Thus, a *toolbox* without tools is still a *toolbox*, which indicates that tools is connected to box via **some other relationship (i.e. container-contained)**. This result is the same for the French *boîte à outils* (i.e. *une boîte à outils sans outils est toujours une boîte à outils*). When applied to compounds that denote a part-whole association, the test produces defective or incomplete readings.

(pp. 196–197, emphasis added)

The case of “honey box” (*beehive*) is parallel to *toolbox*: a beehive without honey is indubitably still a beehive. The distinction Bourque makes is useful, but his conclusion to treat *toolbox* (and thus also “honey box”) as (mere) location seems inadequate. It seems better to bite the bullet and add the relation CONTAINMENT (which even Bourque recognizes as “some other relationship”) to his system, on the grounds that the ability to perceive containment is a fundamental part of our cognitive endowment. The relation is reversible and may be exemplified, following Hatcher (1960: 364), by *orange seed* and *seed orange*.

The other addition made to Bourque’s set of relations was motivated by one of Jespersen’s examples: *sun worship*. Strictly speaking this is not a binominal since *worship* denotes an action, not a thing. However, the scope of Bourque’s classification is noun-noun compounds in general and therefore it should be able to accommodate *sun worship*. It turns out that none of Bourque’s relations are appropriate. Clearly the notion of the sun as some kind of goal is involved, so one might think that Bourque’s SOURCE would do the job, but no amount of tweaking of either the Basic or the Reversed template produces a paraphrase that is acceptable for both *sun worship* and *cane sugar*. This seems to be because goal as a complement of source is not compatible with result. It seems that a new relation is unavoidable, but what to call it, and how to make it sufficiently distinct from SOURCE? The answer is provided by Hatcher, who includes *sun worship* in her category A←B (to be discussed below), pointing out that “the sun is that toward which the worship is *directed*” (see Figure 2; emphasis added). Now, as we have seen, it is frequently the case that the verb used to express the paraphrase can serve in nominalized form as the name of the relation itself (recall ‘possess’ > POSSESSION). The solution to the problem of how to name the new relation is thus given: ‘direct’ > DIRECTION, understood as an asymmetric relation which relates a starting point or origin and an endpoint or goal, and exemplified by *sun worship* and *sales target*, respectively.

Adding such a relation to Bourque’s scheme can be justified on two grounds (over and above the desire to accommodate *sun worship*): firstly, it is very general, and secondly, the ability to conceptualize direction is an important part of the human cognitive endowment. Further research may show that DIRECTION is rarely encountered in binominals, but it may turn out to be more important when synthetic compounds and other complex nominals containing an action-root are considered (as in *sun worship*).

The classification resulting from the modifications to Bourque’s system described in the preceding sections consists of 17 relations, two of them symmetric, and three non-reversible, for a total of 29 (directed) ‘sub-relations’, hence the name “Bourque29”. Documentation for each of these is provided in the Appendix, together with an at-a-glance summary. In the following section we turn to the

high-level classification developed by Anna Granville Hatcher to which Bourque²⁹ will be mapped in §4.

3 Hatcher's high-level classification

3.1 The critique of Jespersen

Hatcher presents her (1960) four-way classification of non-appositional compounds in the form of a critique of Jespersen's (1942) attempt to classify semantic relations. Jespersen concedes that his analysis is incomplete and that there are many compounds which "do not fit in anywhere", but he claims that his failure is simply due to the inherent unclassifiability of his material: "the number of possible logical relations between the two elements is endless" (p. 138); "the analysis of the possible sense-relations can never be exhaustive" (p. 143).

But, says Hatcher,

it all too often happens that scholars in linguistics proclaim a given problem to be insoluble, when they themselves have not worked out the categories necessary for its solution; we should, then, examine the outline offered by Jespersen to see if some of the difficulty he encountered may not be explained by his method of classification. For example, was his set of categories constructed with logical rigor: and, before surrendering to the "difficult" types that he mentions, had he been able, at least, to account for all the "easy" compounds, subdividing these as carefully as his patience and his talent permitted? The subdivision of the obvious may lead to greater understanding of the less obvious, if one is guided by logically consistent criteria. (p. 356)

Thereupon, Hatcher sets about dissecting and reordering Jespersen's system. She starts by listing seven of Jespersen's types, omitting one of the original eight (Similarity) on the grounds that it more properly belongs to "apposition", which she wants to keep separate. Examining each of these in turn, Hatcher notes a lack of careful subdivisions, an absence of any principle of symmetry, and the mixing of two basic criteria, Reference and Relation. Her rearrangement of Jespersen's scheme is depicted in Figure 2.

Hatcher chooses to avoid Reference and to base her new scheme exclusively on Relation, so she starts by separating the first three of Jespersen's types **1–3** (Subject/Object, Place and Time) – all of which are either based on reference or mixed – from types **4–7** (Purpose, Means, Characterizing Feature and Material), all of which are relational. The former are set to one side, and to the latter she adds two relational types found in Mätzner (1860) but absent in Jespersen (**α** *broomstick* and **β** *castor oil*). She then proceeds to reorganize these six relational types into four abstract classes:

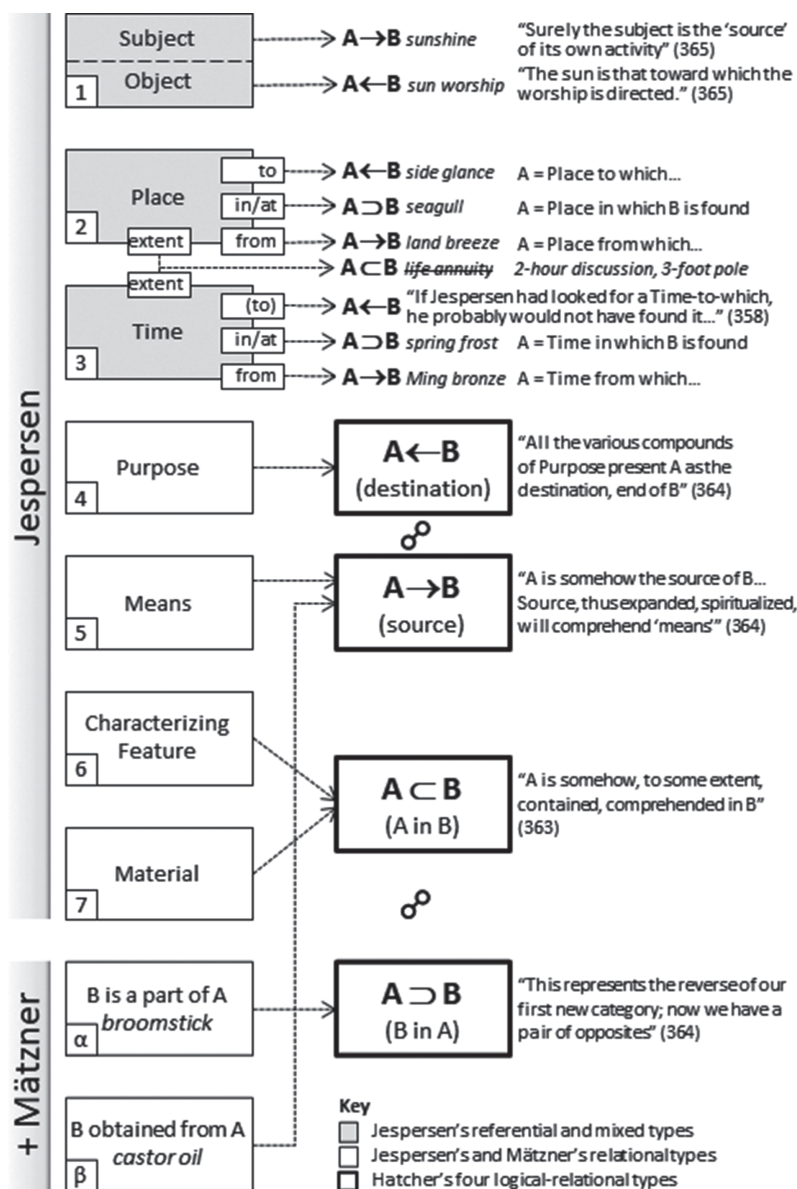


Figure 2: Hatcher's reworking of Jespersen's classification.

- (a) $A \subset B$ “A is contained in B” (notated $\textcircled{A}$ by Hatcher)
- (b) $A \supset B$ “B is contained in A” (notated $\textcircled{B}$ by Hatcher)
- (c) $A \rightarrow B$ “A is the source of B”
- (d) $A \leftarrow B$ “A is the destination of B”

Having reduced the six relational categories of Jespersen/Mätzner to two pairs of mutually exclusive concepts, Hatcher turns her attention to the referential types, in order to see how they might be accommodated in her new scheme. She starts with (2) Place, (3) Time and their subdivisions (*to*, *in/at*, *from* and *extent*), which map neatly into her scheme, as (d), (b), (c) and (a), respectively.

Finally, the two verbal types (1) Subject and (2) Object are “easy”:

Sunshine and *sun worship*, these perfect opposites, fall under $A \rightarrow B$ and $A \leftarrow B$, respectively. Surely the subject is the “source” of its own activity (in putting *sunshine* under $A \rightarrow B$, we are merely adding Agent to Agency); and in *sun-worship* ($A \leftarrow B$), the sun is that toward which the worship is directed.

Thus we see that both the referential and the relational types of Matzner-Jespersen can be included in our two pairs of relational criteria: the static $\textcircled{A}$ and $\textcircled{B}$, and the dynamic $A \rightarrow B$ and $A \leftarrow B$. (p. 365)

Hatcher concludes this part of her analysis by pointing out that the scheme she has developed has two advantages over the one she has just “torn to pieces”. Firstly, it is logically conceived, and therefore neater and more pleasing aesthetically; and secondly, it is far more comprehensive, and thus may “be able to account for all possibilities of determinative, non-appositional compounding in the English language,” which she suggests are surely not “endless” (p. 365–366). At the same time she expresses the hope that her work represents not a “result”, but rather a beginning, and that it will offer “a more spacious framework” within which research dedicated to the proposition that “all compounds are endowed by their creators with the right to belong somewhere” may proceed more profitably and hopefully than before.

3.2 Extending Hatcher’s classification

Hatcher’s work is often cited, but usually dismissed, often on less than scientific grounds. For example, Søgaaard (2005: 320) writes:

such an account is by definition both arbitrary (Bauer 1978; van Santen 1979) and incomplete because of the infinite set of compounding relationships. For illustration, try to place a compound such as *car thief* in [Hatcher’s] four-way typology. Is a car thief a ‘car in a thief’, a ‘thief in a car’, a ‘thief as the goal of a car’ or a ‘thief as the source of a car’?

Unfortunately for Søgaaard the last two paraphrases are incorrect: He has muddled up the order of A and B. The head of the construction (B) is *thief*, not *car*, so these two paraphrases should read: a ‘*car* as the goal of a *thief*’ and a ‘*car* as the source of the *thief*’. With the correct paraphrase, it is obvious that the car is indeed the goal of the thief (i.e. $A \leftarrow B$). Søgaaard’s objection must therefore be rejected.

One researcher who has taken Hatcher seriously is Arnaud (2003; 2016). Arnaud’s work on categorizing the modification relations in French subordinative NN_N compounds is full of interesting observations, examples and discussion. However, in the present context it is noteworthy for the fact that Arnaud first develops his own highly granular classification, and then attempts to map it onto Hatcher’s four-way scheme (which Noailly 1990, also working on French compounds, had arrived at independently).

Arnaud’s classification is based on a database of 949 French binominals of type **cmp** and **jxt**, which he dubs “les composés timbre-poste” (postage stamp compounds). As none of the then-existing taxonomies of semantic relations seemed satisfactory, he decided to start from the data up, applying the principles of cognitive linguistics, “in particular the idea that relations are emergent phenomena which gain psychological existence” (2016: 71). The analysis resulted in a classification with 58 categories, ranging from the highly abstract (e.g. “Non-head is the goal of Head”) to the very precise (such as the subtype of the LOCATION relation “Non-head is a secondary activity taking place in Head”).

Arnaud now proceeds to map his set of 58 empirically derived (low-level) relations to Noailly and Hatcher’s set of four logically derived (high-level) relations. For the most part, this is plain sailing:

In most cases, the fine-grained categories were easy to group under these [high-level relations]. For example, the description in (18) was classified as an instance of (19).

- (18) It is against the effects of Non-head that Head is made/conceived/set up
ex.: *minimum vieillesse* (lit. ‘minimum old-age’, i.e. basic old-age benefits)

- (19) NON-HEAD $\leftarrow$ HEAD
Abstract relation (19) represents the fact that in (18) the denotatum of N_2 is, so to say, aimed at that of N_1 (p. 81).⁶

⁶ Arnaud’s ‘non-head’ and ‘head’ correspond to Hatcher’s A and B. The high-level relation in his (19) is therefore equivalent to her $A \leftarrow B$. It can be useful to think that A stands for Attribute (= modifier, non-head) and B for Base (= head).

Arnaud’s bottom-up deduction thus melds neatly with Hatcher’s top-down induction. Or at least, it almost does. Arnaud experienced difficulties with 12 of his 58 low-level categories that did not map straightforwardly to Hatcher’s four, and he felt obliged to extend Hatcher’s system with four more high-level relations: ANALOG, BE, HEAD SYMB NON-HEAD and NON-HEAD SYMB HEAD.

The frequencies of the four high-level categories are shown in Table 3. Arnaud himself concedes that “[the four new] categories are marginal compared with the initial four,” but he believes they “show that Noailly erred on the side of abstraction (and Hatcher, too, as equivalent English compounds are easily found)” (p. 81).

Table 3: Frequencies of high-level relations in Arnaud (2016).

Relation	Equiv.	Freq.	%
NON-HEAD $\leftarrow$ HEAD	$A \leftarrow B$	428	38.1
((NON-HEAD) HEAD)	$A \subset B$	295	26.3
NON-HEAD $\rightarrow$ HEAD	$A \rightarrow B$	159	14.2
(NON-HEAD (HEAD))	$A \supset B$	126	11.2
ANALOG	–	62	5.5
HEAD SYMB NON-HEAD	–	24	2.1
BE	–	23	2.0
NON-HEAD SYMB HEAD	–	5	0.4

Pepper (2020) examines each of the 12 low-level relations that seemed to Arnaud to justify the creation of his four new high-level relations and shows that all but one of them can in fact be accommodated by Hatcher’s four-way system. For example, the first of Arnaud’s problematic forms, *régime jockey*, denotes a diet that is typical of jockeys. But if A (‘jockey’) typifies (or characterizes) B (‘diet’), then it is a *characterizing feature* of B and therefore belongs, as Figure 2 shows, under Hatcher’s $A \subset B$, “A is somehow, to some extent, contained, comprehended in B”. Thus it turns out, in other words, that Hatcher’s system is broad enough to cater for eleven of the twelve low-level relations that prompted Arnaud to add four new high-level categories.

The single exception, one of four subtypes of ANALOG, is exemplified by the form *brasse papillon* [breast_stroke butterfly] ‘butterfly stroke’, which falls under Arnaud’s low-level category “Non-head names analogically a perceptual characteristic of Head”. Here there can be no doubt that some kind of analogy is at work. But *brasse papillon* is not a non-appositional compound in Hatcher’s terms and therefore falls outside the scope of her 1960 paper.

If we want to extend Hatcher’s scheme to cover appositional compounds, then we do indeed need a new high-level relation. However, ANALOGY may not be the

best term for that relation. Hatcher’s logically defined pair of reversible relations are both based on Contiguity, which is one of Aristotle’s “three principles of remembering”, the others being Similarity and Contrast. In Pepper (2020) I suggest that the relation underlying the types of appositional compound discussed by Hatcher herself in an earlier paper (Hatcher 1952), i.e. species-genus and cross-classification – as well as Arnaud’s *brasse papillon* (and incidentally also coordinative compounds) – is Similarity. This is at about the right level of generality or abstraction as Hatcher’s original two pairs. So her four-way system can be extended to a five-way system consisting of two pairs of asymmetric relations (which Hatcher referred to as ‘static’ and ‘dynamic’) that account for non-appositional compounds, and a fifth, symmetric relation that accounts for appositional compounds.

The extended system (Hatcher5) is summarized in Table 4. Following Bourque, Hatcher’s A and B are replaced with M and H, and machine-readable codes (e.g. *HinM*) have been added as alternatives to notations such as $M \supset H$ or $\textcircled{B}$. Furthermore, Hatcher’s “static” and “dynamic” have been tentatively recast as **CONTAINMENT** and **DIRECTION**, respectively.

Table 4: Revised high-level classification (Hatcher5).

<i>Contiguity-based</i>		
CONTAINMENT (“static”)		
$M \supset H$	<i>HinM</i>	“H is contained in M” (<i>orange seed</i>)
$M \subset H$	<i>MinH</i>	“M is contained in H” (<i>seed orange</i>)
DIRECTION (“dynamic”)		
$M \leftarrow H$	<i>HtoM</i>	“M is the destination of H” (<i>sugar cane</i>)
$M \rightarrow H$	<i>MtoH</i>	“M is the source of H” (<i>cane sugar</i>)
<i>Similarity-based</i>		
SIMILARITY		
$M \cong H$	<i>MisH</i>	“H is similar or identical to M”

4 The Hatcher-Bourque classification

4.1 Description

Mapping the revised Bourque classification to the revised Hatcher system was quite straightforward. Three of the 17 relations are based on similarity in one way or another and thus map to the new relation:

- TAXONOMY equates to what Hatcher (1952) terms the “species-genus” type (e.g. *pumice stone*);
- COORDINATION and SIMILARITY correspond to two subtypes of her “cross-classification” type (exemplified by *fuel oil* and *butterfly table*).

The remaining 14 relations map neatly to Hatcher’s original two pairs of relations as follows:

- CONTAINMENT, POSSESSION, PARTONOMY, LOCATION, TEMPORALITY, COMPOSITION and TOPIC are subtypes of her “static” relations; Basic forms map consistently to *HinM* (“B is contained in A”) and Reversed forms to *MinH* (“A is contained in B”)
- DIRECTION, SOURCE-RESULT, CAUSATION, PRODUCTION, USAGE, FUNCTION and PURPOSE are subtypes of her “dynamic” relations; Basic forms map consistently to *HtoM* (“B is the source of A”) and Reversed forms to *MtoH* (“A is the source of B”).

As Figure 3 shows, the Hatcher-Bourque classification operates at two main levels of granularity, labelled Bourque29 and Hatcher5, respectively. Bourque29 consists of the 17 rather granular, low-level relations, indicated by the codes in the five boxes at the bottom of the diagram. Of these, 12 are reversible, giving a total of 29 (24+5) low-level (directed) ‘sub-relations’ (hence, Bourque29). The low-level relations map to the three schematic, high-level relations of Hatcher5, labelled SIMILARITY, CONTAINMENT and DIRECTION. Of these, the latter two are reversible, for a total of five high-level (directed) ‘sub-relations’ (hence Hatcher5).

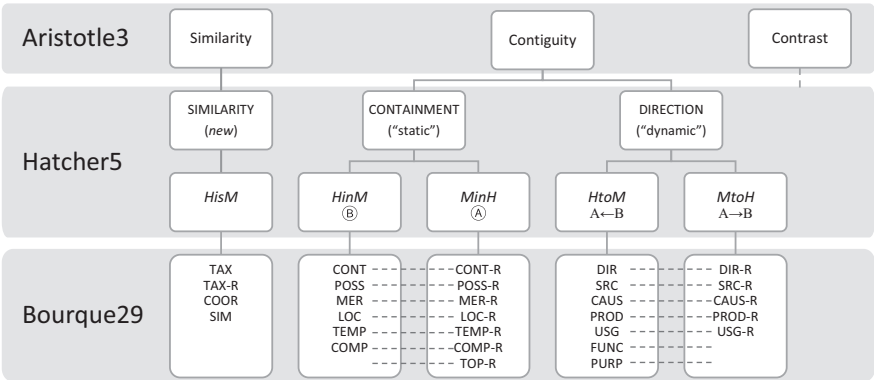


Figure 3: The Hatcher-Bourque classification as a hierarchy.

The relations of CONTAINMENT and DIRECTION are both based on Contiguity, one of Aristotle's three principles of memory, while Similarity constitutes another of those principles (Koch 2001: 1143). The third principle, Contrast, appears to play only a very minor role in binominal word-formation and has not yet been investigated in detail. It is therefore not part of this initial version of Hatcher-Bourque. However, examples do exist, for example Mandarin 东西 *dōng.xī* [east.west] 'thing' (Ceccagno & Scalise 2006: 238). This justifies including a placeholder in Figure 3.

The complete classification is documented in the Appendix in the form of descriptions of each individual low-level relation and an at-a-glance summary table (Table 7).

4.2 The Bourquifier: A piece of cake

Classifying large numbers of binominals can be a daunting and error-prone task, even with a well-documented classification that includes test frames and examples. In order to simplify the task and reduce the risk of errors, an Excel application called the *Bourquifier* has been created (Pepper 2021). This tool is designed to assist the analyst, not to replace her. The way it works is by the analyst typing the head and modifier (and optionally the binominal itself) into the relevant cells, upon which all 29 templates are automatically populated. These can then be scanned in a matter of seconds to find the most appropriate relation.

The interface of the *Bourquifier* (Figure 4) shows the 17 low-level relations of Bourque29 listed under the heading *Relation*, and the roles associated with each of them in the adjoining column. (Note that for the two symmetric relations, COORDINATION and SIMILARITY, the two roles are the same.) These relations are grouped according to the three high-level relations of Hatcher5: SIMILARITY, CONTAINMENT and DIRECTION.

To the right of the column headed *Roles* the interface is divided into two sections, for Basic and Reversed forms of the relation, respectively. Each section consists of four columns: one for the B29 code (e.g. TAX-R), one for the corresponding H5 code (e.g. *MisH*), one for the template (e.g. “(an) M is a kind of H”) and one for the example (e.g. *oak tree*). For symmetric and non-reversible relations, one section is blank.

Figure 5 shows how the *Bourquifier* is used to analyse a specific example, here *sunburn*, which it may be recalled from §2.2 was erroneously chosen by Bourque to exemplify his Basic CAUSE relation (see Table 2). The populated templates in the *Bourquifier* make it very clear that *sunburn* actually belongs under the Reversed

Bourquifier 4 (Hatcher-Bourque 29/5V2)				Binominal (B)		Modifier (M)		Head (H)		Relation
				B		M		H		ANY
H3	Relation	Roles	B29	H5	Basic template	Example	B29	H5	Reversed template	Example
SIMILARITY	TAXONOMY	SUPERTYPE SUBTYPE	TAX	H5M	(an) H is a kind of M	bear cub	TAX-R	H5M	(an) M is a kind of H	oak tree
	COORDINATION	COORDINATE COORDINATE	COORD		(an) H that is also (an) M	boy king			(symmetric)	
	SIMILARITY	LIKENESS LIKENESS	SIM		(an) H that is similar to (an) M	kidney bean				
CONTAINMENT	CONTAINMENT	CONTAINER CONTENTS	CONT	H5M	(an) H that is contained in (an) M	orange seed	CONT-R	H5M	(an) H that contains (an) M	seed orange
	POSSESSION	POSSESSOR POSSESSEESUM	POSS		(an) H that is possessed by (an) M	family estate	POSS-R		(an) H that possesses (an) M	career girl
	PARTONOMY	WHOLE PART	PART		(an) H that is part of (an) M	car motor	PART-R		(an) H that (an) M is part of	motor car
	LOCATION	LOCATION LOCATED	LOC		(an) H located at/near/in (an) M	house music	LOC-R		(an) H that (an) M is located at/near/in	music hall
	TEMPORALITY	TIME ACTIVITY	TEMP		(an) H that occurs at/during (an) M	summer job	TEMP-R		(an) H at/during which (an) M occurs	golf season
	COMPOSITION	COMPOSITE MATERIAL	COMP		(an) H that (an) M is composed of	cube sugar	COMP-R		(an) H that is composed of M	sugar cube
DIRECTION	TOPIC	ENTITY TOPIC	COMP	H5M	(non-reversible)		TOP-R	H5M	(an) H that is about (an) M	history book
	DIRECTION	GOAL ORIGIN	DIR		(an) H whose goal is (an) M	sun worship	DIR-R		(an) H that is the goal of (an) M	sales target
	SOURCE-RESULT	RESULT SOURCE	SRC		(an) H that is a source of (an) M	sugar cane	SRC-R		(an) H whose source is (an) M	cane sugar
	CAUSATION	EFFECT CAUSE	CAUS		(an) H that causes (an) M	tear gas	CAUS-R		(an) H that (an) M causes	sunburn
	PRODUCTION	PRODUCT PRODUCER	PROD		(an) H that produces (an) M	song bird	PROD-R		(an) H that (an) M produces	birdsong
	USAGE	USER USED	USG		(an) H that (an) M uses	lamp oil	USG-R		(an) H that uses (an) M	oil lamp
	FUNCTION	FUNCTION ENTITY	FUNC		(an) H that serves as (an) M	buffer state			(non-reversible)	
	PURPOSE	PURPOSE ENTITY	PURP		(an) H intended for (an) M	animal doctor				

Figure 4: The Bourquifier interface.

Bourquifier 4 (Hatcher-Bourque 29/5V2)				Binominal (B)		Modifier (M)		Head (H)		Relation
				B		M		H		CAUS-R
H3	Relation	Roles	B29	H5	Basic template	Example	B29	H5	Reversed template	Example
SIMILARITY	TAXONOMY	SUPERTYPE SUBTYPE	TAX	H5M	(a) burn is a kind of sun	bear cub	TAX-R	H5M	(a) sun is a kind of burn	oak tree
	COORDINATION	COORDINATE COORDINATE	COORD		(a) burn that is also (a) sun	boy king			(symmetric)	
	SIMILARITY	LIKENESS LIKENESS	SIM		(a) burn that is similar to (a) sun	kidney bean				
CONTAINMENT	CONTAINMENT	CONTAINER CONTENTS	CONT	H5M	(a) burn that is contained in (a) sun	orange seed	CONT-R	H5M	(a) burn that contains (a) sun	seed orange
	POSSESSION	POSSESSOR POSSESSEESUM	POSS		(a) burn that is possessed by (a) sun	family estate	POSS-R		(a) burn that possesses (a) sun	career girl
	PARTONOMY	WHOLE PART	PART		(a) burn that is part of (a) sun	car motor	PART-R		(a) burn that (a) sun is part of	motor car
	LOCATION	LOCATION LOCATED	LOC		(a) burn located at/near/in (a) sun	house music	LOC-R		(a) burn that (a) sun is located at/near/in	music hall
	TEMPORALITY	TIME ACTIVITY	TEMP		(a) burn that occurs at/during (a) sun	summer job	TEMP-R		(a) burn at/during which (a) sun occurs	golf season
	COMPOSITION	COMPOSITE MATERIAL	COMP		(a) burn that (a) sun is composed of	cube sugar	COMP-R		(a) burn that is composed of sun	sugar cube
DIRECTION	TOPIC	ENTITY TOPIC	COMP	H5M	(non-reversible)		TOP-R	H5M	(a) burn that is about (a) sun	history book
	DIRECTION	GOAL ORIGIN	DIR		(a) burn whose goal is (a) sun	sun worship	DIR-R		(a) burn that is the goal of (a) sun	sales target
	SOURCE-RESULT	RESULT SOURCE	SRC		(a) burn that is a source of (a) sun	sugar cane	SRC-R		(a) burn whose source is (a) sun	cane sugar
	CAUSATION	EFFECT CAUSE	CAUS		(a) burn that causes (a) sun	tear gas	CAUS-R		(a) burn that (a) sun causes	sunburn
	PRODUCTION	PRODUCT PRODUCER	PROD		(a) burn that produces (a) sun	song bird	PROD-R		(a) burn that (a) sun produces	birdsong
	USAGE	USER USED	USG		(a) burn that (a) sun uses	lamp oil	USG-R		(a) burn that uses (a) sun	oil lamp
	FUNCTION	FUNCTION ENTITY	FUNC		(a) burn that serves as (a) sun	buffer state			(non-reversible)	
	PURPOSE	PURPOSE ENTITY	PURP		(a) burn intended for (a) sun	animal doctor				

Figure 5: The Bourquifier ('sunburn').

form of the relation, with the paraphrase “(a) burn that (a) sun causes”. (Note that the highlighting on CAUS-R is a result of the analyst typing the code into the red box in the top right-hand corner; it does not happen automatically.)

It is worth noting at this point that sometimes more than one paraphrase will apply to a single binominal. For example, *motor car* (Figure 6) may be analysed as

- “a car that contains a motor” (CONT-R: Reversed CONTAINMENT),
- “a car that a motor is part of” (PART-R: Reversed PARTONOMY), or
- “a car that a motor is located at/near/in” (LOC-R: Reversed LOCATION).

Bourquifier 4				Binominal (B)		Modifier (M)	Head (H)		Relation
(Hatcher-Bourque 23/5/2)				motor car		motor	car		PART-R
H3	Relation	Roles	B29	H5	Basic template	Example	B29	H5	Reversed template
SIMILARITY	TAXONOMY	SUPERTYPE SUBTYPE	TAX	HiM	(a) car is a kind of motor	bear cub	TAX-R	HiM	(a) motor is a kind of car
	COORDINATION	COORDINATE COORDINATE	COORD		(a) car that is also (a) motor	boy king			
	SIMILARITY	LIKENESS LIKENESS	SIM		(a) car that is similar to (a) motor	kidney bean			(symmetric)
CONTAINMENT	CONTAINMENT	CONTAINER CONTENTS	CONT		(a) car that is contained in (a) motor	orange seed	CONT-R		(a) car that contains (a) motor
	POSSESSION	POSSESSOR POSSESSEES	POSS		(a) car that is possessed by (a) motor	family estate	POSS-R		(a) car that possesses (a) motor
	PARTONOMY	WHOLE PART	PART	HiM	(a) car that is part of (a) motor	car motor	PART-R	HiM	(a) car that (a) motor is part of
	LOCATION	LOCATION LOCATED	LOC		(a) car located at/near/in (a) motor	house music	LOC-R		(a) car that (a) motor is located at/near
	TEMPORALITY	TIME ACTIVITY	TEMP		(a) car that occurs at/during (a) motor	summer job	TEMP-R		(a) car at/during which (a) motor occurs
	COMPOSITION	COMPOSITE MATERIAL	COMP		(a) car that (a) motor is composed of	cube sugar	COMP-R		(a) car that is composed of motor
DIRECTION	TOPIC	ENTITY TOPIC			(non-reversible)		TOP-R		(a) car that is about (a) motor
	DIRECTION	GOAL ORIGIN	DIR		(a) car whose goal is (a) motor	sun worship	DIR-R		(a) car that is the goal of (a) motor
	SOURCE-RESULT	RESULT SOURCE	SRC		(a) car that is a source of (a) motor	sugar cane	SRC-R		(a) car whose source is (a) motor
	CAUSATION	EFFECT CAUSE	CAUS		(a) car that causes (a) motor	tear gas	CAUS-R		(a) car that (a) motor causes
	PRODUCTION	PRODUCT PRODUCER	PROD	HiM	(a) car that produces (a) motor	song bird	PROD-R	HiM	(a) car that (a) motor produces
	USAGE	USER USED	USG		(a) car that (a) motor uses	lamp oil	USG-R		(a) car that uses (a) motor
	FUNCTION	FUNCTION ENTITY	FUNC		(a) car that serves as (a) motor	buffer state			
	PURPOSE	PURPOSE ENTITY	PURP		(a) car intended for (a) motor	animal doctor			(non-reversible)

Figure 6: The Bourquifier (‘motor car’).

When every candidate relation maps to the same high-level relation (as is the case here, since all three relations map to *MinH*), we have a simple case of overlap between very similar relations. In such cases, either relation may be used, but the more specific relation (here, PARTONOMY) is usually to be preferred. However, sometimes the candidate relations map to different high-level relations – as would be the case if “a car that uses a motor” (USG-R: Reversed USAGE) were considered an appropriate paraphrase for *motor car* – since this relation maps to *MtoH*. In such cases the combination of concepts can be considered to be “doubly motivated”; i.e. the combination motor + car is motivated both by the PARTONOMY relation and by the USAGE relation. There is nothing untoward about this, since there is no reason to believe that every combination of concepts should be motivated by a single relation.⁷

7 Those that have tried the *Bourquifier* have found it very helpful. Readers can see for themselves that analysing a binominal is a piece of cake by taking part in the Hatcher-Bourque Cake Challenge. Simply download the *Bourquifier* (see the URL in the References) and use it to analyse the seven examples given in §1.1. Send me your results and I will buy you coffee and cake next time we meet. The results of my own analysis are given at the end of this chapter, but don’t change your results to fit these. The point of the exercise is to see how much inter-annotator agreement is achieved using the *Bourquifier*.

5 Frequency of semantic relations

The Hatcher-Bourque classification was developed as part of a broader investigation into the typology and semantics of binominal lexemes (Pepper 2020), and it was used to classify 3,738 binominals from 106 languages denoting 100 different concepts.⁸ Only 83 of these binominals (2.2%) resisted classification.⁹ 79 of them were simply unanalysable, either because of a cranberry morpheme, as in Chakali [CLI] *nebi.kan̄kawal* [finger.??] ‘thumb’, or because the motivation is veiled by unfamiliar beliefs or cultural practices, as in Takia [TBC] *tamol sos* [man Derris_root] ‘widower’. Four binominals use a numeral modifier to denote a day of the week (e.g. Iraqw [IRK] *deelór tām* [day:of three] ‘Wednesday’), for which no appropriate relation exists. However, since such cases are more properly regarded as instances of property modification rather than object modification, they are outside the scope of Hatcher-Bourque. The remaining 3,650 binominals were easily analysed using the *Bourquifier*. The resulting data lends itself to an analysis of the relative frequency of semantic relations cross-linguistically.

It is not unreasonable to surmise that the frequency with which different semantic relations are used to motivate the combination of concepts in binominal word-formation could provide insights into the way in which humans conceptualize the world. This is a topic which has hardly been addressed in the typological literature at all; to my knowledge, the only researcher to even approach the question from a cross-linguistic perspective is Bauer (2001), who has the following to say:

In a detailed survey of just three languages, Bauer (1978: 147) points out that underlying semantic relationships of location appear to be the most common relationships in those languages. The same is true with the sample [of 36 languages] discussed here. Compounds in which the head is the location of the entity denoted in the modifier (e.g. English *furniture store*) or where the head denotes an entity located at the modifier (e.g. English *bone cancer*) are the types most frequently illustrated or commented on for the languages in my sample across all areas. The next most frequent type to be illustrated is the type where the head is made from the material in the modifier (e.g. English *sandcastle*). Other meanings are illustrated or commented on far more sporadically. While this does not show that other meanings are not also in common use, it does suggest that compounds may be used prototypically to indicate location or source (especially if ‘made from’, ‘made by’, ‘belonging to’ and ‘coming from’ are all interpreted as sources).

⁸ Sources for all material mentioned in this section can be found in Pepper (2020).

⁹ In addition five entries were considered to be incorrect, in the sense that the form registered in the database does not express the intended meaning. For example, the Yaqui word *muumu jo'ara* [bee house] almost certainly denotes a beehive and not beeswax, as stated in the source. Such cases could have been analysed in their own terms (in this case as BASIC POSSESSION), but instead they were simply excluded, in order not to distort the analysis of individual meanings.

With Hatcher-Bourque, Bauer's three examples (*bone cancer*, *furniture store* and *sandcastle*) are classified as Basic LOCATION ("a cancer located at/near/in a bone"), Reversed LOCATION ("a store that furniture is located at/near/in") and Reversed COMPOSITION ("a castle composed of sand"). We can now use the binominals data to test Bauer's conjecture. The frequencies of the Bourque29 low-level relations are investigated in §5.1, and those of the Hatcher5 high-level relations in §5.2.

5.1 Frequency of low-level relations

The overall frequency of low-level semantic relations in the database, shown in Figure 7, can be summarized in the following scale:

PART >> PURP > COOR > LOC > COMP-R, POSS > USG-R > TEMP > . . .

By far the most frequent relation is one that Bauer does not even mention: PART. This is the Basic PARTONOMY (or whole) relation, in which an entity is modified by the whole of which it is part, as in *car motor*. The quite extreme frequency of this relation may be due to the large number of binominals in the database that denote body parts, which tend to be based on this relation (as in *eyelid*). For this reason, Figure 7 also shows the frequencies when body parts are excluded entirely. Apart from the greatly reduced frequency of PART and a slightly reduced frequency for LOC (the location relation) the differences are minimal. So while it may be the case that the present data overstate the prevalence of PART, it is clearly one of the most important relations, and probably more frequent than LOC and LOC-R combined.

Bauer's suggestion that the next most frequent type is the material relation (Reversed COMPOSITION), e.g. *sandcastle*, is also not supported by the data, which put it at joint fifth in terms of overall frequency. Instead, the next most frequent relation is purpose, also not mentioned by Bauer. As we will see below, this relation is especially prevalent in binominals that belong to the domain *Modern World* and/or denote entities that fall into the semantic type Advanced technology (or concept).

The third most frequent relation is COORDINATION. In the binominals data, it is mostly found in items that denote animates of a certain age (Hawaiian [HAW] *kao keiki* [goat child] KID), gender (Mbyá Guaraní [GUN] *kavaju kunha* [horse woman] MARE), or both (Ket [KET] *qīm.duīl* [woman.child] GIRL). However, it should be borne in mind that the set of meanings on which these data are based was designed to exclude many kinds of coordination relation (such as Vietnam-

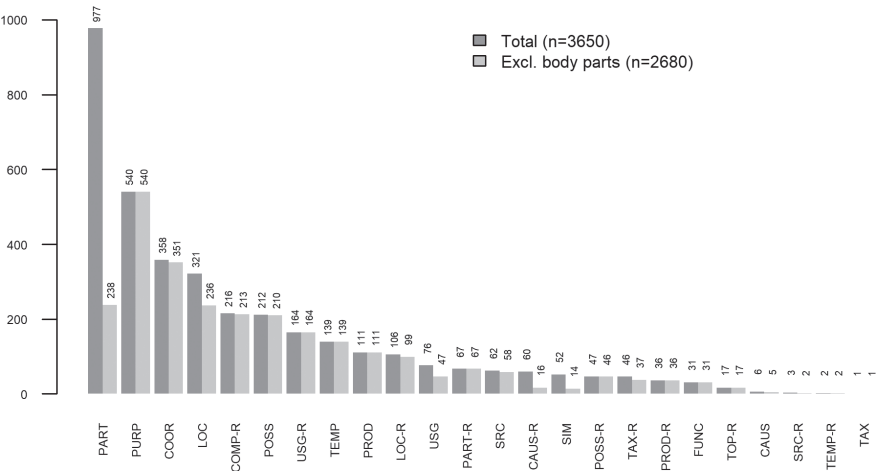


Figure 7: Overall frequency of low-level semantic relations.

ese [VIE] *bố mẹ* [father mother] PARENTS), so the prevalence of species-attribute combinations cannot be taken as fully representative.

The location relation (Basic LOCATION), found when words denoting eye and water are combined to denote TEAR, is only the fourth most frequent. Together with its inverse, the located relation, for example Hupdë [JUP] *yǝʰ mɔy* [medicine house] HOSPITAL, it is found in 428 binominals, i.e. 12% of the data. Thus Bauer's suggestion that this is the most common kind of relation is clearly unsupported.

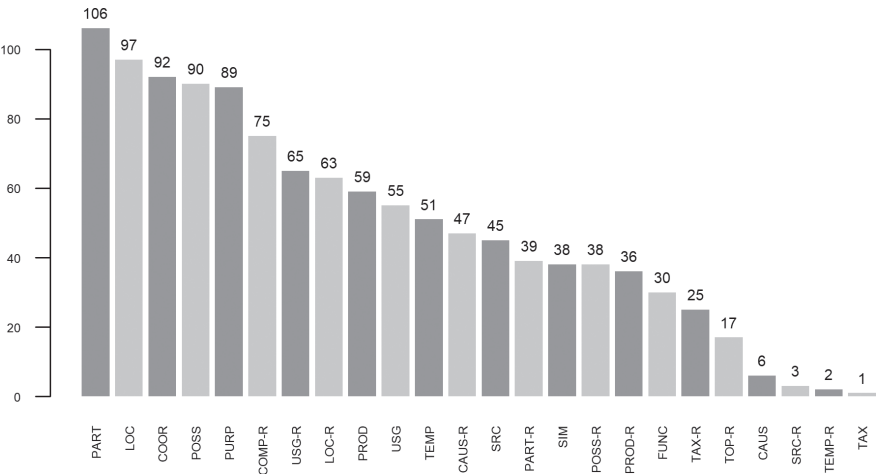


Figure 8: Number of languages that exhibit a particular relation.

We can also look at relations in terms of the number of languages in which each relation is attested (Figure 8). The frequency scale here is:

PART > LOC > COOR, POSS, PURP > COMP-R > USG-R, LOC-R, PROD > ...

The same six relations predominate in both scales, albeit with slightly different rankings. Note that the composite relation (COMP, e.g. *cube sugar*) is not attested at all in the database. Note also the infrequency of a further four – CAUS, SRC-R, TEMP-R and TAX – in which the modifier expresses the effect (*tear gas*), the source (*cane sugar*), the (temporally located) activity (*golf season*) and the supertype (*bear cub*).

The distribution across meanings (Figure 9) shows a generally similar scale, but now with the TAX-R relation displaying far greater prominence. USG now appears among the top six, with COMP-R and POSS relegated to joint 9th and 11th place:

PART > COOR, PURP, TAX-R, LOC > USG, SIM, USG-R, POSS, LOC-R, COMP-R...

This suggests that while the subtype relation, TAX-R (*oak tree*) is not especially common, it is rather versatile in terms of the range of meanings that it can express. Conversely, while the material (COMP-R) and possessor (POSS) relations are rather frequent, their scope of application is relatively limited. It is also worth noting that of the 46 binominals that exhibit the subtype relation (Figure 7), 18 employ the **der** strategy. In many cases, the gloss indicates an (apparently redundant) nominalizer or diminutive affixed to a root whose meaning is the same as that of the derived form, as in Lithuanian [LIT] *spen.elis* [nipple.DIM] NIPPLE OR TEAT.

Overall, the data indicate that the most frequent low-level semantic relations cross-linguistically, at least as far as binominal lexemes are concerned, are as shown in Table 5.

Table 5: Most frequent low-level semantic relations.

Relation	Modifier role	Template	example
PART	<u>whole</u>	(an) H that is part of (an) M	<i>car motor</i>
PURP	<u>purpose</u>	(an) H intended for (an) M	<i>animal doctor</i>
COOR	<u>coordinand</u>	(an) H that is also (an) M	<i>boy king</i>
LOC	<u>location</u>	an H that (an) M is located at/near/in	<i>house music</i>
POSS	<u>possessor</u>	(an) H that (an) M possesses	<i>family estate</i>
COMP-R	<u>material</u>	(an) H composed of (an) M	<i>sugar cube</i>

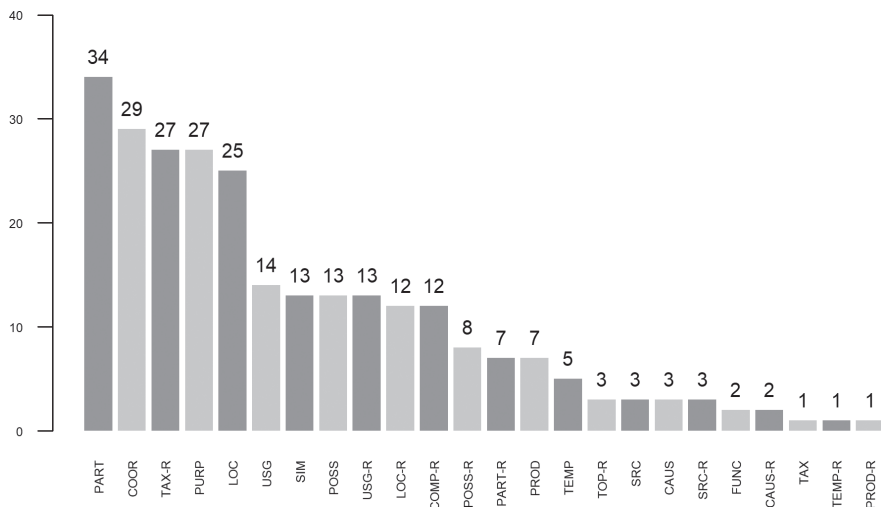


Figure 9: Number of meanings that exhibit a particular relation.

Figure 10 shows how many of the nine morphosyntactic strategies (as defined in the binominal typology described in Pepper, a, this volume) are used to express each kind of relation. Comparison with the overall frequency scale extracted from Figure 7 (above) shows that the most frequent relations can be expressed by any one of the nine binominal types. This very strongly suggests that there is no overall correlation – at the cross-linguistic level – between morphosyntactic strategies and semantic relations. However, this should not be taken to mean that there is no such correlation at the level of individual languages. On the contrary, studies such as Pepper (2010) show that semantic relation can be an important explanatory factor in the study of intra-linguistic competition between binominal strategies.

As the data become sparser, the number of strategies associated with each relation declines; thus, at the lower end of the scale, we find TEMP-R, SRC-R and TAX, each of which is expressed by just one or two strategies. However, since each of these three relations is represented in the database by just two or three exemplars, this does not constitute evidence against the lack of overall correlation.

The frequency of different relations varies according to the semantic type of the referent. Figure 11 shows the proportional distribution of the six most common relations – PART, PURP, COOR, LOC, COMP-R and POSS – across seven semantic types. The results for Animal, Natural phenomenon and Location should be approached with caution, since these semantic types represent only 7, 5 and 12 of the 100 meanings, respectively, but the variation across the other four types is striking.

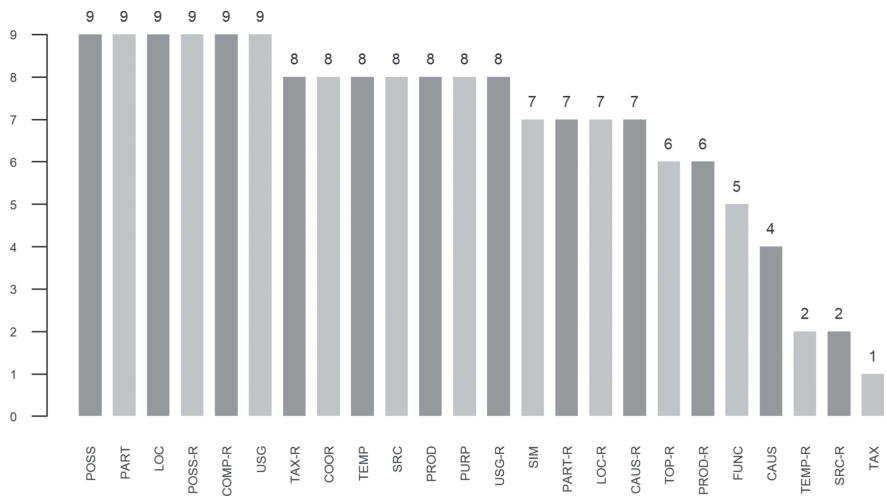


Figure 10: Number of binominal pairs that exhibit each relation.

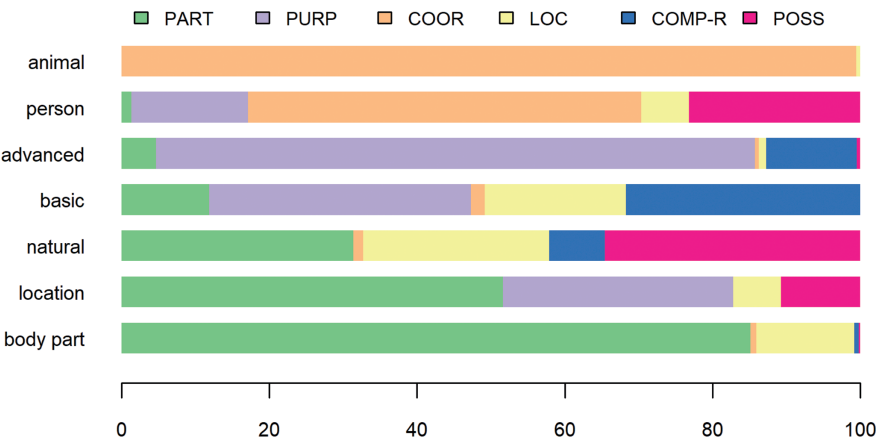


Figure 11: Low-level relations and semantic types.

In binominals denoting Body parts the whole relation (PART) accounts for 85% of the data; the only significant alternative is located (LOC), which is the preferred relation for naming bodily substances, such as EARWAX and TEAR. On the other hand, PART is rarely used to denote an Advanced technology (or concept), such as BICYCLE PUMP, KEYWORD or RAILWAY; instead, the purpose relation predominates, accounting for over 80% of the data, with material (COMP-R) the most frequently used alternative (as in many words for RAILWAY, which is often conceptualised as a road composed of iron). In short, there is a strong tendency to name

(secondary) body parts/fluids in terms of the (primary) body parts they are a part of/located at, and to name advanced concepts in terms of either their intended function or the material they are made of.

The semantic type Basic technology (or concept) is more mixed: as with Advanced technology (or concept), purpose and material are the most widespread relations, but the two are now equally frequent; however, in contradistinction to the latter, the whole (PART) and location (LOC) relations are also quite frequent. These are also the most widely used relations for Natural phenomena – together with possessor (POSS), which expresses the relation between a spider and its web, or bees and their hive, as well as phenomena viewed as belonging to some supernatural being, such as Ket *Albara kân* ‘Milky Way, lit. Alba’s hunting trail’ and Assamese [ASM] *ramdheni* ‘rainbow, lit. Lord Rama’s bow’.

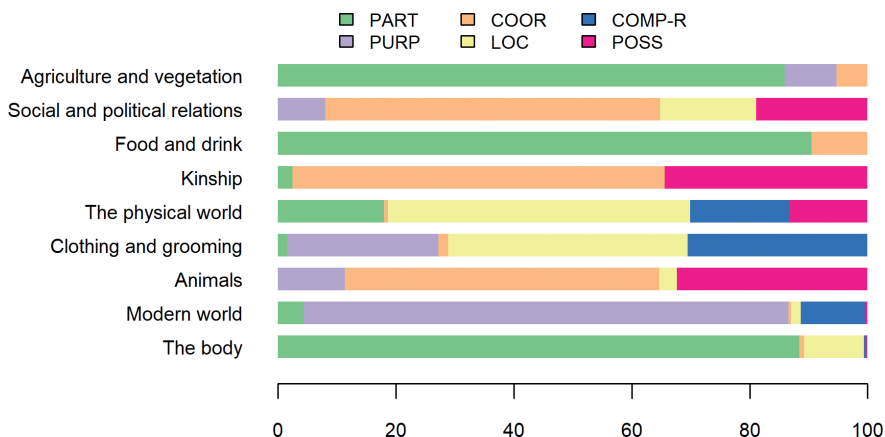


Figure 12: Low-level relations and semantic fields.

A similar variation is found across semantic fields. Figure 12 shows the frequency of the six most common semantic relations across the nine most frequent semantic fields. We note again that PART plays the dominant role in The body, but also in Agriculture and vegetation and Food and drink; and, as expected, Modern world is dominated by the purpose relation. We see also that the patterning in Animals and Kinship is remarkably similar: binominals in these fields have an overwhelming preference for either COOR or POSS. The latter is also widely used in Social and political relations. Finally, the location relation (LOC) that Bauer assumed to be most widespread is in fact largely confined to the fields of The physical world and Clothing and grooming.

5.2 Frequency of high-level relations

We turn now from the low-level semantic relations of Bourque29 to the high-level relations of Hatcher5. For ease of reference, Table 6 provides a summary of the five high-level relations and the 29 low-level relations that map to them. (Note that CONTAINMENT and DIRECTION were not used to annotate the contents of the binominals database and therefore do not figure in the statistics of the preceding section.) As for the low-level relations, the terms in the role column will sometimes be used in the following in order to simplify the discussion. They are in effect shorthand labels for the Hatcher5 ‘sub-relations’.

Table 6: Summary of mappings from high- and low-level.

Hatcher5	Modifier role	Bourque29
<i>MisH</i>	N/A	TAX-R, TAX, COOR, SIM
<i>HinM</i>	<u>container</u>	CONT, POSS, PART, LOC, TEMP, COMP, TOP
<i>MinH</i>	<u>contents</u>	CONT-R, POSS-R, PART-R, LOC-R, TEMP-R, COMP-R
<i>HtoM</i>	<u>goal</u>	DIR, SRC, CAUS, PROD, USG, FUNC, PURP
<i>MtoH</i>	<u>origin</u>	DIR-R, SRC-R, CAUS-R, PROD-R, USG-R

The first four plots in the previous section showed how the low-level relations distribute across the database as a whole (with and without body parts), and across languages, meanings and morphosyntactic strategies.

Figure 13 provides similar information for the high-level relations. Predictably, the information content is considerably reduced; on the other hand, the categories are much more balanced and therefore more amenable to statistical analysis.

The first thing to note is that every one of the nine morphosyntactic strategies is attested in the data as expressing each of the five high-level relations (plot d); this provides additional evidence that there is no overall, cross-linguistic correlation between morphosyntactic strategies and semantic relations. (Again, this does not mean that such correlations do not exist within individual languages.)

The high-level container relation *HinM* (Hatcher’s “B is contained in A”) accounts for nearly half of the data (a). This comes as no surprise, given that this relation subsumes PART. If body parts are excluded it has roughly the same frequency as the goal relation *HtoM* (Hatcher’s “A is the destination of B”), which subsumes the rather frequent purpose relation, among others. With body parts included, the overall scale is as follows (>> denotes very significantly more frequent than; > denotes significantly more frequent than):

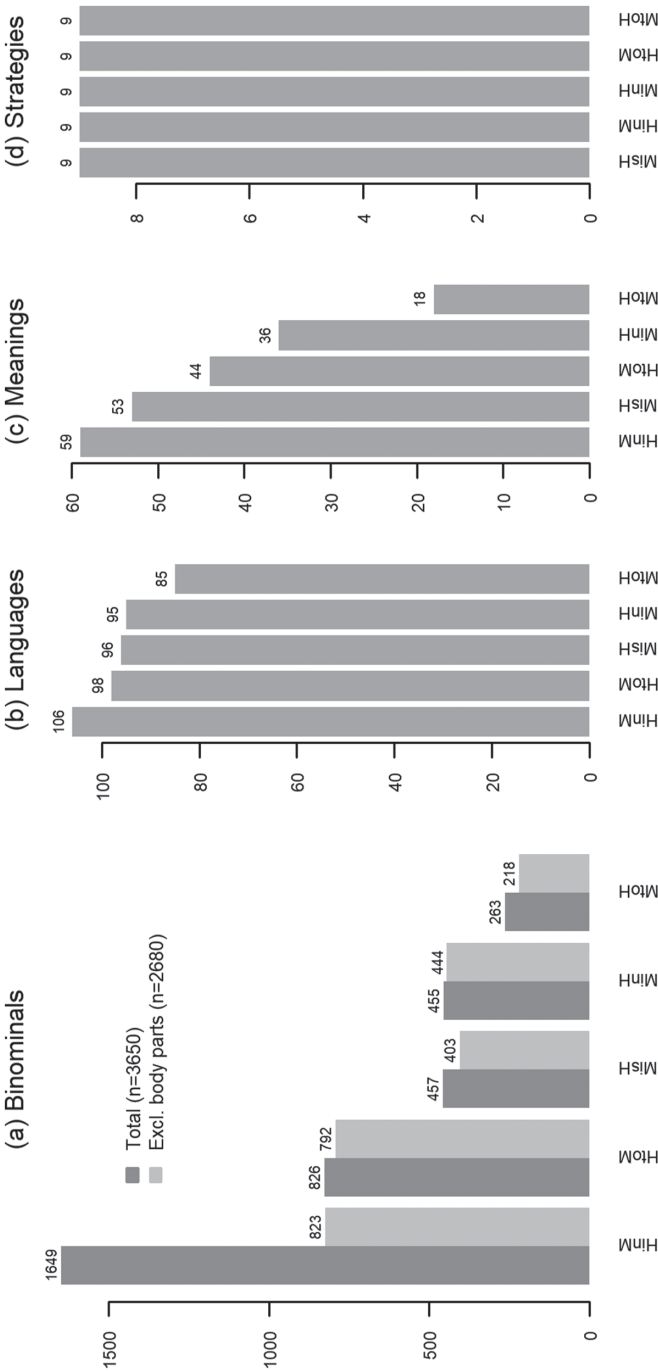


Figure 13: High-level relations across binominals, languages, meanings and strategies.

HinM >> *HtoM* > *MisH* *MinH* > *MtoH*

With body parts excluded, the scale is

HinM, *HtoM* > *MisH*, *MinH* > *MtoH*

The two most frequent low-level relations (*HinM* and *HtoM*) account for two-thirds of the data and thus suggest a pronounced tendency for a complex meaning to be conceptualized in terms of either its container or its goal – both of which should be interpreted in Hatcher’s very broad sense.

Plot (b) tells us that *HinM* is ubiquitous, occurring in every language in the sample. However, the other four low-level relations are also widespread across languages and they are probably also ubiquitous. The fact that they are not attested in every language is almost certainly due to the paucity of data for some languages: it would be highly unlikely that a language that is represented by fewer than, say, ten data points¹⁰ would exhibit all five high-level relations.

The distribution of relations across meanings (c) shows a scale similar to the two preceding ones –

HinM > *HtoM* > *MisH* > *MinH* > *MtoH*

– but the values are more spread out: *HinM* is less dominant, while *MisH*, the similarity-based relation added to Hatcher’s original four is higher up the scale (in the sense that it is significantly more widespread across meanings than *MinH*). This reflects what was referred to above as the versatility of the subtype relation (TAX-R). More worthy of mention, though, is the fact that none of the high-level relations appears suited for conceptualizing anything like the full range of meanings. Even *HinM*, which is found in every language and accounts for over 45% of all binominals in the database, is used with only just over half of the 100 meanings: in other words, there are limits to the versatility of conceptualizations that are based on how an entity is (in the broadest sense) “contained”.

With regard to semantic types, Figure 14 shows clearly that the container relation (*HinM*) is central to the conceptualization of (secondary) Body parts and also important for concepts that express Location or that denote Basic technologies (or concepts) and for entities in the Natural world. On the other hand, it is marginal to the conceptualization of Persons and of almost no use when it comes

¹⁰ There are five of these in the database: Gurindji [GUE], Puyuma [PYU], Selice Romani [RMC], Datooga [TCC] and Tuwari [TWW].

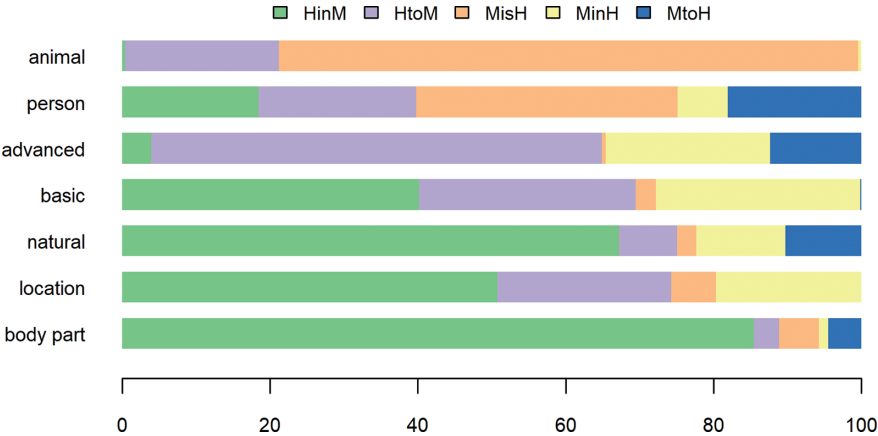


Figure 14: High-level relations and semantic types.

to Animals and Advanced technologies (or concepts). With the semantic types Animal and Person (and only those) the similarity-based *HinM* relation is most important, whereas conceptualizations that are goal-oriented – indicated by the *HtoM* relation – are most frequent with Advanced technologies (and concepts), but also encountered with other semantic types (albeit only rarely with Body parts and Natural phenomena).

Conceptualization of an entity in terms of its contents (*MinH*) is considerably less common than the inverse and never the dominant form; it is found most often with semantic types that denote Basic and Advanced technologies (and concepts) and Locations, rarely with Body parts and never with Animals. As for origin-based conceptualizations, they are mostly found with Persons (in particular, professions), Natural phenomena, and Advanced technologies (and concepts).

Similar patterns emerge with respect to semantic fields (Figure 15, the high-level equivalent of Figure 12). Whereas the low-level plot highlights similarities between Animals and Kinship, the new one reveals additional commonalities, in particular between The body, The physical world and Food and drink. In all of these, the container relation (*HinM*) predominates: there is a tendency for conceptualizations where (to quote Hatcher 1960: 363–364) the target concept, B, “is somehow, to some extent, contained, comprehended in” the modifying concept, A.

In sum, and referring back to the notion of roles, we see that the container (*HinM*) is particularly important for The body, Food and drink and The physical world; the goal (*HtoM*) for the Modern world; similarity (*MisH*) for Kinship and Animals; contents (*MinH*) for Clothing and grooming and for Social and political relations; and origin (*MtoH*) for Agriculture and vegetation.

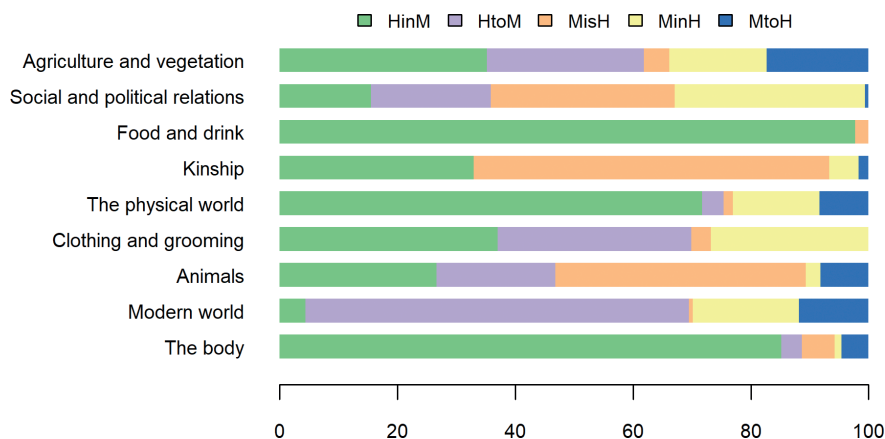


Figure 15: High-level relations and semantic fields.

5.3 Discussion

The analysis of the data provides insights into the ways in which humans tend to conceptualize the world. It suggests, *contra* Bauer, that PARTONOMY and PURPOSE are far more widespread, and thus more important, than the LOCATION relation. Of the two types of PARTONOMY – Basic (PART) and Reversed (PART-R) – the former is far more frequent than the latter, which indicates that the conceptualization of a complex meaning is much more likely to involve modification by the whole (or, more generally, the container) than modification by the parts (or, more generally, the contents). The Basic PARTONOMY relation (PART) occurs most frequently with body parts and in the semantic field of agriculture and vegetation. It can express about one third of the 100 meanings used in this survey; it is found in all 106 languages of the sample; and it can be expressed using any one of the nine nominal modification strategies.

Bauer's suggestion that the next most frequent type is where the head is made from the material in the modifier is also not supported by the data: both PURPOSE and COORDINATION are much more common than COMPOSITION. The purpose relation is most often encountered in the semantic field Modern world to denote advanced technological concepts; it only occurs in 89 of the 106 languages, no doubt because some of the languages in the sample do not have words for concepts of that kind; significantly, the only morphosyntactic strategy that does not occur with this relation is the classifier strategy, **cls**, but this is also the most sparsely populated of all strategies. COORDINATION is used primarily to denote animates of a certain age, gender or both; it is therefore unsurprising that

it occurs mostly in the domains of kinship, animals, agriculture and vegetation. Cases such as these account for over 90% of binominals that exhibit this relation, and once again, every morphosyntactic strategy is attested in the data.

The Basic LOCATION relation is the fourth most frequent type overall and occurs three times as often as its reverse; in other words, it is more usual to conceptualize an object in terms of where it is located than what is located at, near or in it. It is found in almost all of the languages of the sample (97 out of 106) and can be expressed by any of the nine strategies. It is most often encountered in the fields of the natural world and basic technologies and concepts.

The other fairly frequent relations are those of possessor (Basic POSSESSION) and material (Reversed COMPOSITION). The range of meanings that can be expressed by these two is limited: only 12% in each case; all the same, they can be expressed by any strategy. On the other hand, the reversed form of POSSESSION is uncommon, and the Basic form of COMPOSITION does not occur in the data at all.

Apart from the latter, every one of the 25 relations used for annotation was found in the data, but some were very rare, in particular those involving modification by an effect (e.g. *tear gas*, CAUS), a source (*cane sugar*, SRC-R), a temporal activity (*golf season*, TEMP-R), or a supertype (*bear cub*, TAX). While these are fairly peripheral in binominals, they may be more common in other types of compounds, for example those in which the head or the modifier is an action-morph rather than a thing-morph (see Pepper, this volume, a for the precise definition of binominal used in the present study).

The data for the low-level relations suggests that there is no overall correlation between morphosyntactic strategy and semantic relation: many relations are expressed by every strategy, most are expressed by almost every strategy, and those that are expressed by just a few strategies are those where the data is sparse. This impression is confirmed by the analysis of high-level relations: every one of the five relations of Hatcher5 are attested with every one of the nine morphosyntactic strategies, so we can state quite categorically that there is no such overall correlation. It is thus not the case some strategies are used to express some relations, while other strategies are used for other relations.

However, while this applies cross-linguistically, it does not mean that there are no such correlations within individual languages. In fact, the opposite is the case: As I showed in Pepper (2010), the Cameroonian language Nizaa uses left-headed and right-headed compounds for two distinct sets of relations. Zúñiga (2014) reports something similar for Mapudungun [ARN], as does Atoyebi (2010) for Oko [OKS]. Bourque himself (p. 253) compares N N and N à N binominals in French and shows that the two constructions have very different profiles (for example, PURPOSE and USE account for 48% of all French N à N binominals in his database, but only 13% of his N N binominals). Some of the contributions in this

volume start to address this issue for other languages, but there is much work to be done. That work would be much more productive if researchers were to adopt the same classification system, and that is the purpose of Hatcher-Bourque.

6 Summary and further work

In this chapter I started out by providing a brief overview of previous studies on semantic relations in compounding. I then described in some detail the systems developed by Bourque and Hatcher and how they were harnessed in the present study. Bourque's classification was revised and extended with two new relations, CONTAINMENT and DIRECTION, for a total of 29 relations, 12 reversible and five unidirectional. Hatcher's classification was also revised – by the addition of a fifth high-level relation, SIMILARITY – in order to extend its coverage to appositional as well as non-appositional compounds. The two revised classifications were then unified to create the two-tiered Hatcher-Bourque classification, and an Excel-based tool called the *Bourquifier* was developed to assist in the slippery task of classifying individual binominals. Both Hatcher-Bourque and the *Bourquifier* are offered to the research community in order to promote collaboration in the field of semantic relations.¹¹

It is important to state that the current version of Hatcher-Bourque (29/5/v1) is a work-in-progress. It needs to be tested against more data from more languages. It may still need refining, through improved examples and templates, and perhaps even the addition of more relations. Certainly CONTRAST needs to be fleshed out, and COORDINATION could be subdivided to better handle the variation currently covered by this category. Perhaps it should be possible to distinguish between partial and full COMPOSITION? If so, this can be done by increasing the granularity. Could one conceive of logical subdivisions between the two layers of Bourque29 and Hatcher5, such as grouping SOURCE-RESULT, CAUSATION and PRODUCTION (on the one hand) and USE, FUNCTION and PURPOSE (on the other) within Hatcher's 'dynamic' pairing of DIRECTION-based relations? And why *exactly* are some asymmetric relations apparently non-reversible?

¹¹ Hatcher-Bourque Cake Challenge (§4.2). The results of my analysis of the seven cake examples in §1.1 are as follows: *chocolate cake*: material (COMP-R); *birthday cake*: purpose (PURP); *coffee cake*: UK material (COMP-R) / US purpose (PURP); *marble cake*: likeness (SIM); *layer cake*: material (COMP-R); *cup cake*: container (CONT); *urinal cake*: location (LOC).

Appendix: Documentation for Hatcher-Bourque 29/5/v1

This appendix documents version 1 of the Hatcher-Bourque 29/5 classification.¹² In the following presentation, the 17 low-level relations are grouped according to the three high-level relations, SIMILARITY, CONTAINMENT and DIRECTION. Reference in square brackets, e.g. [5.2.2.1], are to the extended discussion of the relation (possibly under another name) in Bourque (2014).

1 Similarity-based relations

The similarity-based relations are those found in what Hatcher (1952) calls “appositional” compounds. Hatcher identified two basic types, “species-genus” (e.g. *pumice stone*) and “cross-classification” (e.g. *fuel oil, butterfly table*). The former corresponds to TAXONOMY, and the latter to COORDINATION and SIMILARITY, respectively. These all map to the similarity-based high-level relation, *MisH*.

TAXONOMY (supertype / subtype)

Basic	TAX	<i>MisH</i>	“an H is a kind of M”	<i>bear cub</i>
Reversed	TAX-R	<i>MisH</i>	“an M is a kind of H”	<i>oak tree</i>

The relation between a type (e.g., *tree*) and one of its subtypes (e.g., *oak*). Both constituents satisfy the ISA test: an oak tree is an oak, and an oak tree is a tree. In addition, and crucially, every oak is a tree. In the Basic form, the superordinate concept is denoted by the modifier (*bear* in *bear cub*), and in the Reversed form by the head (*tree* in *oak tree*). The Reversed form of this relation is sometimes called the species-genus relation, and compounds that exhibit it are sometimes called pleonastic, epexegetic or subsumptive. [5.2.2.1 HYPONYMY; inverted]

¹² See Pepper (2020) for the earlier version. Changes between the two are documented in Pepper (2021).

COORDINATION (coordinand)

Symmetric COOR *MisH* “an H that is also an M” *boy king*

When this relation pertains, both constituents (*boy* and *king*) satisfy the ISA test: a boy king is both a boy and a king. However, there is no type-subtype relation between the two: it is not the case that every boy is a king, and neither is every king a boy. This is the crucial difference between the COORDINATION and TAXONOMY relations. [5.2.2.2]

SIMILARITY (likeness)

Symmetric SIM *MisH* “an H that is similar to an M” *kidney bean*

In this relation the modifying concept has some characteristic feature in common with the referent. In the case of *kidney bean*, it is shape: a kidney bean is a bean shaped like a kidney. [5.2.2.3]

2 Containment-based relations

The containment-based relations are finer-grained subtypes of Hatcher’s high-level relations, “A is somehow, to some extent, contained, comprehended in B” (*MinH*), and its inverse, “B is somehow, to some extent, contained, comprehended in A” (*HinM*).

CONTAINMENT (container / contents)

Basic CONT *HinM* “an H that is contained in an M” *orange seed*
 Reversed CONT-R *MinH* “an H that contains an M” *seed orange*

The relation between a container and its contents: the seed is contained in the orange and the orange contains the seed. In *orange seed*, the modifier denotes the container, whereas in *seed orange*, the modifier denotes the contents. (See Pepper 2020: 226–227 for further discussion.)

POSSESSION (possessor / possessum)

Basic	POSS	<i>HinM</i>	“an H that is possessed by an M”	<i>family estate</i>
Reversed	POSS-R	<i>MinH</i>	“an H that possesses an M”	<i>career girl</i>

The relation between a possessor and a possessum, both in the specific sense of ownership (*family estate*) and the more general sense of belonging (*career girl*). [5.2.2.5; inverted]

PARTONOMY (whole / part)

Basic	PART	<i>HinM</i>	“an H that is part of an M”	<i>car motor</i>
Reversed	PART-R	<i>MinH</i>	“an H that an M is part of”	<i>motor car</i>

The relation between a whole and one of its parts. A motor can be specified in terms of the car of which it is a part (*car motor*), and a car can be specified in terms of one of its most salient parts (*motor car*). [5.2.2.6 PART]

LOCATION (location / located)

Basic	LOC	<i>HinM</i>	“an H located at/near/in an M”	<i>house music</i>
Reversed	LOC-R	<i>MinH</i>	“an H that M is located at/near/in”	<i>music hall</i>

The relation between an entity or activity (the thing located) and its location. A music hall is a hall in which (a certain kind of) music is (or was) performed. The origin of the term ‘house music’ is unclear, but it is likely that ‘house’ refers to the location in which the music was either created or performed. This relation may be restricted to spatial locations; relations involving a temporal location use TEMPORALITY. [5.2.2.7]

TEMPORALITY (time / activity)

Basic	TEMP	<i>HinM</i>	“an H that occurs at/during an M”	<i>summer job</i>
Reversed	TEMP-R	<i>MinH</i>	“an H at/during which M occurs”	<i>golf season</i>

The relation between an entity or activity and the time period during which it occurs, i.e. its temporal location. A summer job is something performed during

the summer; a golf season is the time period during which golf is pursued. [5.2.2.12 TIME]

COMPOSITION (composite / material)

Basic	COMP	<i>HinM</i>	“an H that an M is composed of”	<i>cube sugar</i>
Reversed	COMP-R	<i>MinH</i>	“an H composed of an M”	<i>sugar cube</i>

The relation between a composite entity and the material of which it is composed. The relation inherent in *cube sugar* and *sugar cube* is one and the same (the cube is composed of sugar). The difference is that the one denotes the material (sugar), the other, the composite object (cube). [5.2.2.8; inverted]

TOPIC (entity / topic)

Basic				
Reversed	TOP	<i>MinH</i>	“an H that is about an M”	<i>history book</i>

The relation between an entity or event and the topic that it is “about”: a history book is a book that is about history. An alternative template – “an H that is concerned with an M” – may produce a more felicitous paraphrase, as in the case of *history department*: a department that is concerned with history. [5.2.2.11]

3 Direction-based relations

The direction-based relations in this section are finer-grained subtypes of Hatcher’s high-level relations “A is somehow the source of B” (*MtoH*) and its inverse “B is somehow the source of A” (*HtoM*).

DIRECTION (goal / origin)

Basic	DIR	<i>HtoM</i>	“an H whose goal is an M”	<i>sun worship</i>
Reversed	DIR-R	<i>MtoH</i>	“an H that is the goal of an M”	<i>sales target</i>

The relation between a point of origin (usually an activity) and its goal. In *sun worship*, the sun is the goal towards which the worship is directed, and a *sales target* is that towards which a sales activity is directed. (See Pepper 2020: 227–228 for further discussion.)

SOURCE-RESULT (result / source)

Basic	SRC	<i>HtoM</i>	“an H that is a source of an M”	<i>sugar cane</i>
Reversed	SRC-R	<i>MtoH</i>	“an H whose source is an M”	<i>cane sugar</i>

The relation between a source and a result – in a general sense that does not involve either causation or production; in *sugar cane*, while the cane is the source of the sugar, it cannot felicitously be said to cause or produce it. [5.2.2.9 SOURCE; inverted]

CAUSATION (effect / cause)

Basic	CAUS	<i>HtoM</i>	“an H that causes an M”	<i>tear gas</i>
Reversed	CAUS-R	<i>MtoH</i>	“an H that an M causes”	<i>sunburn</i>

The relation between a cause and an effect. Tear gas is a gas that causes tears; sunburn is a burn that is caused by the sun. [5.2.2.10 CAUSE]

PRODUCTION (product / producer)

Basic	PROD	<i>HtoM</i>	“an H that produces an M”	<i>song bird</i>
Reversed	PROD-R	<i>MtoH</i>	“an H that an M produces”	<i>birdsong</i>

The relation between a product and its producer. Both *song bird* and *birdsong* involve the production of song by a bird, but whereas in the former, the modifier denotes the product, in the latter it denotes the producer. [5.2.2.10]

USAGE (used / user)

Basic	USG	<i>HtoM</i>	“an H that an M uses”	<i>lamp oil</i>
Reversed	USG-R	<i>MtoH</i>	“an H that uses an M”	<i>oil lamp</i>

The relation between something that is “used” and the entity (“user”) that uses it. An oil lamp uses oil, and its oil is used by the lamp. In *lamp oil* the modifier denotes the user, while the modifier of *oil lamp* denotes the thing used. [5.2.2.13 USE; inverted]

FUNCTION (function / entity)

Basic FUNC *HtoM* “an H that serves as an M” *buffer state*
 Reversed

The relation between an entity and its function: a buffer state is a state that serves as a buffer. Unlike PURPOSE (below), this relation does not involve any element of intentionality. Despite being asymmetric, it does not appear to be reversible. [5.2.2.4]

PURPOSE (purpose / entity)

Basic PURP *HtoM* “an H that is intended for an M” *animal doctor*
 Reversed

The relation between an entity and its purpose: an animal doctor is a doctor whose skills are directed towards animals. Unlike FUNCTION (above), this relation involves an element of intentionality. Despite being asymmetric, it does not appear to be reversible. [5.2.2.14]

Table 7: The Hatcher-Bourque classification.

Bourque29	B29	H5	Template	Example
TAXONOMY <u>supertype, subtype</u>	TAX-R	MisH	an <u>M</u> is a kind of <u>H</u>	<i>oak tree</i>
	TAX	MisH	an <u>H</u> is a kind of M	<i>bear cub</i>
COORDINATION <u>coordinand, coordinand</u>	COOR	MisH	an <u>H</u> that is also an M	<i>boy king</i>
SIMILARITY <u>likeness, likeness</u>	SIM	MisH	an <u>H</u> that is similar to M	<i>kidney bean</i>
CONTAINMENT <u>container, contents</u>	CONT	HinM	an <u>H</u> that is contained in an M	<i>orange seed</i>
	CONT-R	MinH	an <u>H</u> that contains an M	<i>seed orange</i>
POSSESSION <u>possessor, possessum</u>	POSS	HinM	an <u>H</u> that is possessed by an M	<i>family estate</i>
	POSS-R	MinH	an <u>H</u> that possesses an M	<i>career girl</i>
PARTONOMY <u>whole, part</u>	PART	HinM	an <u>H</u> that is part of an M	<i>car motor</i>
	PART-R	MinH	an <u>H</u> that an M is part of	<i>motor car</i>
LOCATION <u>location, located</u>	LOC	HinM	an <u>H</u> located at/near/in an M	<i>house music</i>
	LOC-R	MinH	an <u>H</u> that M is located at/near/in	<i>music hall</i>
TEMPORALITY <u>time, event</u>	TEMP	HinM	an <u>H</u> that occurs at/during an M	<i>summer job</i>
	TEMP-R	MinH	an <u>H</u> at/during which M occurs	<i>golf season</i>
COMPOSITION <u>composite, material</u>	COMP	HinM	an <u>H</u> that an M is composed of	<i>cube sugar</i>
	COMP-R	MinH	an <u>H</u> composed of an M	<i>sugar cube</i>
TOPIC <u>entity, topic</u>	TOP-R	MinH	an <u>H</u> that is about an M	<i>history book</i>
DIRECTION <u>goal, origin</u>	DIR	HtoM	an <u>H</u> whose goal is an M	<i>sun worship</i>
	DIR-R	MtoH	an <u>H</u> that is the goal of an M	<i>sales target</i>
SOURCE <u>result, source</u>	SRC	HtoM	an <u>H</u> that is a source of an M	<i>sugar cane</i>
	SRC-R	MtoH	an <u>H</u> whose source is an M	<i>cane sugar</i>
CAUSATION <u>effect, cause</u>	CAUS	HtoM	an <u>H</u> that causes an M	<i>tear gas</i>
	CAUS-R	MtoH	an <u>H</u> that an M causes	<i>sunburn</i>
PRODUCTION <u>product, producer</u>	PROD	HtoM	an <u>H</u> that produces an M	<i>song bird</i>
	PROD-R	MtoH	an <u>H</u> that an M produces	<i>birdsong</i>
USAGE <u>user, used</u>	USG	HtoM	an <u>H</u> that an M uses	<i>lamp oil</i>
	USG-R	MtoH	an <u>H</u> that uses an M	<i>oil lamp</i>
FUNCTION <u>function, entity</u>	FUNC	HtoM	an <u>H</u> that serves as an M	<i>buffer state</i>
PURPOSE <u>purpose, entity</u>	PURP	HtoM	an <u>H</u> that is intended for an M	<i>animal doctor</i>

References

- Adams, Valerie. 1973. *An introduction to modern English word-formation*. London: Longman.
- Arnaud, Pierre J.L. 2003. *Les composés timbre-poste*. Lyon: Presses Universitaires de Lyon.
- Arnaud, Pierre J.L. 2016. Categorizing the modification relations in French relational subordinative [NN]_n compounds. In Pius ten Hacken (ed.), *The semantics of compounding*, 71–93. Cambridge: Cambridge University Press.
- Atoyebi, Joseph Dele. 2010. *A reference grammar of Oko: A West Benue-Congo language of North-Central Nigeria*. Rüdiger Köppe.
- Baron, Irène & Michael Herslund. 2001. Semantics of the verb HAVE. In Irène Baron, Michael Herslund & Finn Sørensen (eds.), *Dimensions of possession*, 85–98. Amsterdam: John Benjamins.
- Bauer, Laurie. 1978. *The grammar of nominal compounding: With special reference to Danish, English and French*. Odense: Odense University Press.
- Bauer, Laurie. 1979. On the need for pragmatics in the study of nominal compounding. *Journal of Pragmatics* 3(1). 45–50.
- Bauer, Laurie. 2001. Compounding. In Martin Haspelmath, Ekkehard König, Wolfgang Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, 695–707. Berlin: Mouton de Gruyter.
- Bergsten, Nils. 1911. *A study on compound substantives in English*. Uppsala University PhD dissertation.
- Bourque, Yves. 2014. *Toward a typology of semantic transparency: The case of French compounds*. University of Toronto PhD dissertation.
- Butnariu, Cristina, Su Nam Kim, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz & Tony Veale. 2009. SemEval-2010 Task 9: The interpretation of noun compounds using paraphrasing verbs and prepositions. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (DEW '09)*, 100–105. Stroudsburg, PA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1621969.1621987>.
- Carr, Charles Telford. 1939. *Nominal compounds in Germanic*. London: University of Oxford Doctoral dissertation.
- Ceccagno, Antonella & Sergio Scalise. 2006. Classification, structure and headedness of Chinese compounds. *Lingue e linguaggio* V(2). 233–260.
- Downing, Pamela. 1977. On the creation and use of English compound nouns. *Language* 53(4). 810–842.
- Eiesland, Eli-Anne. 2016. *The semantics of Norwegian noun-noun compounds: A corpus-based study*. University of Oslo PhD dissertation.
- Girju, Roxana, Dan Moldovan, Marta Tatu & Daniel Antohe. 2005. On the semantics of noun compounds. *Computer Speech & Language* 19(4). 479–496 (Special Issue on Multiword Expression). <https://doi.org/10.1016/j.csl.2005.02.006>.
- Girju, Roxana, Preslav Nakov, Vivi Nastase, Stan Szpakowicz, Peter Turney & Deniz Yuret. 2009. Classification of semantic relations between nominals. *Language Resources and Evaluation* 43(2). 105–121.
- Grimm, Jacob. 1826. *Deutsche Grammatik*: 2. Göttingen: Dieterichsche Buchhandlung.
- Hatcher, Anna Granville. 1952. Modern appositional compounds of inanimate reference. *American Speech* 27(1). 3–15.

- Hatcher, Anna Granville. 1960. An introduction to the analysis of English noun compounds. *Word* 16(3). 356–373.
- Jackendoff, Ray. 2009. Compounding in the Parallel Architecture and Conceptual Semantics. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, 105–128. Oxford: Oxford University Press.
- Jackendoff, Ray. 2010. The ecology of English noun-noun compounds. In Ray Jackendoff, *Meaning and the lexicon: The parallel architecture 1975–2010*, 413–451. Oxford: Oxford University Press.
- Jackendoff, Ray. 2016. English noun-noun compounds in Conceptual Semantics. In Pius ten Hacken (ed.), *The semantics of compounding*, 15–53. Cambridge: Cambridge University Press.
- Jespersen, Otto. 1942. *A modern English grammar on historical principles. Part 6: Morphology*. London: George Allen and Unwin.
- Koch, Peter. 2001. Lexical typology from a cognitive and linguistic point of view. In Martin Haspelmath, Ekkehard König, Wolfgang Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: an international handbook*, 1142–1178. Berlin: Mouton de Gruyter.
- Lauer, Mark. 1995. *Designing statistical language learners: Experiments on compound nouns*. Macquarie University PhD dissertation.
- Lees, Robert B. 1960. *The grammar of English nominalizations*. Bloomington: Indiana University.
- Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.
- Marchand, Hans. 1960. *The categories and types of present-day English word-formation*. Wiesbaden: Harrassowitz.
- Mätzner, Eduard. 1860. *Englische Grammatik. Vol. 1 Die Lehre vom Worte*. Berlin: Weidmannsche Buchhandlung.
- Moldovan, Dan, Adriana Badulescu, Marta Tatu, Daniel Antohe & Roxana Girju. 2004. Models for the semantic classification of noun phrases. In *Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics*, 60–67. Association for Computational Linguistics.
- Nakov, Preslav. 2013. On the interpretation of noun compounds: Syntax, semantics, and entailment. *Natural Language Engineering* 19(3). 291–330.
- Noailly, Michèle. 1990. *Le substantif épithète*. Paris: Presses Universitaires de France.
- Ó Séaghdha, Diarmuid. 2008. *Learning compound noun semantics*. University of Cambridge, Computer Laboratory.
- Pepper, Steve. 2010. *Nominal compounding in Nizaa: A cognitive perspective*. SOAS University of London Master's thesis. <https://www.academia.edu/4237937>.
- Pepper, Steve. 2020. *The typology and semantics of binominal lexemes: Noun-noun compounds and their functional equivalents*. Oslo: University of Oslo PhD dissertation. <https://www.academia.edu/42935602>.
- Pepper, Steve. 2021. *The Bourquifier: An application for applying the Hatcher-Bourque classification*. MS Excel. <https://www.academia.edu/83122396>.
- Pepper, Steve. This volume, a. Defining and typologizing binominal lexemes. In Steve Pepper, Francesca Masini & Simone Mattioli (eds.), *Binominal lexemes in cross-linguistic perspective*. Berlin: Mouton de Gruyter.
- Rosario, Barbara & Marti A. Hearst. 2001. Classifying the semantic relations in noun compounds via a domain-specific lexical hierarchy. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*, 82–90.

- Ryder, Mary Ellen. 1994. *Ordered chaos: The interpretation of English noun-noun compounds*. Berkeley: University of California Press.
- Santen, A. van. 1979. Een nieuw voorstel voor een transformationelle behandeling van composita en bepaalde adjectief-substantief combinaties. *Spectator* 9. 240–262.
- Schäfer, Martin. 2018. *The semantic transparency of English compound nouns*. Berlin: Language Science Press.
- Shoben, Edward J. 1991. Predicating and nonpredicating combinations. In Paula J. Schwanenflugel (ed.), *Psychology of word meanings*, 117–135. Hillsdale, NJ: Psychology Press.
- Søgaard, Anders. 2005. Compounding theories and linguistic diversity. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Linguistic diversity and language theories*, 319–337. Amsterdam: John Benjamins.
- Szubert, Andrzej. 2012. *Zur internen Semantik der substantivischen Komposita im Dänischen*. Wydawnictwo Naukowe UAM.
- Toquero, Luis Miguel. 2018. *The semantics of Spanish compounding: An analysis of NN compounds in the Parallel Architecture*. West Virginia University MA thesis.
- Tratz, Stephen & Eduard Hovy. 2010. A taxonomy, dataset, and classifier for automatic noun compound interpretation. In *48th Annual Meeting of the Association for Computational Linguistics*, 678–687. Uppsala: Association for Computational Linguistics.
- Vanderwende, Lucy. 1994. Algorithm for automatic interpretation of noun sequences. In *Proceedings of the 15th conference on Computational linguistics*, vol. 2, 782–788. Association for Computational Linguistics.
- Warren, Beatrice. 1978. *Semantic patterns of noun-noun compounds*. Gothenburg: Acta Universitatis Gothoburgensis.
- Zúñiga, Fernando. 2014. Nominal compounds in Mapudungun. In Swintha Danielsen, Katja Hannss & Fernando Zúñiga (eds.), *Word formation in South American languages*, 11–31. Amsterdam: John Benjamins.

