No amount of experimentation can ever prove me right; a single experiment can prove me wrong.	Albert Einstein
Those who make many species are the 'splitters', and those who make few are the 'lumpers'.	Charles Darwin

1.1 What is a category and how do we find one?

Linguistic descriptions of natural languages typically make reference to grammatical categories (*c*). This monograph addresses three questions: What are grammatical categories? How do we identify them? And are they universal?

What is labeled a grammatical category in individual grammars is not a homogeneous class. Specifically, it includes (but is not limited to) words, morphemes (meaningful units that may be smaller than words), features (that may or may not be associated with an overt expression), as well as certain construction types. These are exemplified below on the basis of categories that are attested in English.¹ To refer to this heterogeneous set of categorizable entities, I use the term *Unit of Language* (UoL).

- (1) Categorizable Units of Language
 - a. Words: Determiners, Complementizers, Auxiliaries, ...
 - b. Morphemes: POSSESSIVE, PROGRESSIVE, ...
 - c. Features: TENSE, NUMBER, CASE, ...
 - d. Clause-types: IMPERATIVE, SUBJUNCTIVE, ...

We talk about a *category* when we can make generalizations over the distribution of a whole set of UoLs. For example, if we know that a word belongs to a certain category c, then we automatically know the distribution of this word. Crucially, this distribution cannot be determined based on either the meaning

¹ The classification in terms of words, morphemes, features, and clause-types is meant for illustrative purpose only. These notions, as we shall see, have no theoretical status.

or the sound of the word. But where does this categorial identity come from? Is it part of a universal repository of categories that is part of our genetic endowment, i.e., part of a universal grammar? Or does it emerge as a matter of language use?

To explore this question it is essential to know whether all languages make use of the same categories, and if not, what the range of variation is. But how can we tell whether categories are universal and if they are universal, how do we identify them? Answering these questions is not a trivial task.

To appreciate its complexity, consider first a more modest question: how do we identify the categories of individual languages? Since its categorial identity determines the morphological and syntactic distribution of a given UoL, we can use distributional criteria to identify categories. For example, we identify a word as an auxiliary if it precedes a main verb (2), if it inflects for tense (3) and subject agreement (4), and if it participates in subject–auxiliary inversion (5).

- (2) a. Edward has blown the whistle.b. Edward is blowing the whistle.
- (3) a. Edward had blown the whistle.b. Edward was blowing the whistle.
- (4) a. They have blown the whistle.b. They were blowing the whistle.
- (5) a. *Has* Edward blown the whistle?b. *Is* Edward blowing the whistle?

Based on these diagnostics, we can identify *have* and *be* as belonging to the category auxiliary, as in (6), where π stands for the representation of its phonetic form.

(6) a. c:AUXILIARY = π :haveb. c:AUXILIARY = π :be

The diagnostic tests for individual categories are always language-specific. For example, not all languages make use of an inflectional category tense. Similarly, subject–auxiliary inversion is not universally attested. Hence neither tense inflection nor subject–auxiliary inversion can function as universal diagnostics for a category auxiliary.

But if criterial diagnostics for categories are language-specific, how do we discover universal categories? In order to identify universal categories, we need universal diagnostics.

And in fact, as I will now show, there are certain formal characteristics of grammatical categories that cut across language-specific patterns in that they go beyond individual sound-meaning associations. These characteristics concern the way UoLs relate to their interpretation. What we observe is that the categorial identity c of a given UoL (i.e., its distribution) plays a critical role in the way this UoL relates to its interpretation. That is, the relation between a UoL and its interpretation is mediated by its categorial identity c. This suggests that the existence of c is a linguistic reality.

1.1.1 Patterns of multifunctionality

To see how c mediates the relation between a UoL and its interpretation, consider again the UoLs *have* and *be*. Based on language-specific criteria, they are classified as auxiliaries, as we have seen above. However, there are also occurrences of these particular forms (*have* and *be*) that do not satisfy the criterial diagnostics for auxiliaries. For example, in (7), they do not precede a main verb: in fact they behave themselves like main verbs. And in (8), we observe that only *be* but not *have* undergoes subject–auxiliary inversion.

- (7) a. Edward has courage.b. Edward is the whistle-blower we have been waiting for.
- (8) a. Does Edward have courage?b. Is Edward the whistle-blower we have been waiting for?

What we observe here is that, both *have* and *be* can be used in two different ways: as main verbs and as auxiliaries. In their use as main verbs, their meaning can roughly be characterized as indicating possession and identity, respectively. This is illustrated in (9), where Σ represents their substantive content² and the curly brackets around π and Σ reflect the fact that they create a unit in the form of an unordered set.

(9) a. $c:VERB = \{\pi: have, \Sigma: possession\}$ b. $c:VERB = \{\pi: be, \Sigma: identity\}$

In their use as auxiliaries, their meaning is hard to pin down. They are not associated with any kind of substantive content, at least not in any obvious way, as indicated by ? in (10). Instead they serve a grammatical function: to form complex tenses.

² I use the term *substantive content* to refer to the type of conceptual content whose interpretation is independent of the linguistic context.

UoL 2 Interpretation 2

Figure 1.1 Multifunctionality as homophony

(10) a. c:AUXILIARY = { π :have, Σ :?} b. c:AUXILIARY = { π :be, Σ :?}

A commonly held view in light of this multifunctionality is to treat it as an instance of accidental homophony (at least synchronically), such that there are two distinct UoLs each associated with a different interpretation, as in Figure 1.1.

But this type of multifunctionality is pervasive across unrelated languages. For example in Halkomelem (Salish) the UoLs π : *i* and π : *li* can be used as lexical verbs (*be here* and *be there*) and as auxiliaries. Consequently they can co-occur within the same sentence, as illustrated in (11)–(12).

(11)	a. $l = chap \ ole i$? AUX-2PL PRT be.here	
	'You folks are here, eh?'	
	·	Galloway 2009: 100
	b. <i>li í the-l tàl?</i> AUX be.here DET.FEM-1SG.POSS mother 'Is my mother in?'	·
		Galloway 2009: 100
(12)	a. $i:-lh = tsel$ li . AUX-PST-1SG.S be.there 'I was there '	
	i was diele.	Galloway 2009: 103
	b. $\mathbf{l}\mathbf{i}$ - $lh = a = chxw$ $\mathbf{l}\mathbf{i}$. AUX-PST-Q-2SG.S be.there 'Were you there?'	
		Galloway 2009: 217

Thus, the Halkomelem auxiliaries show the same pattern of multifunctionality as those of English, as illustrated in (13)–(14).

- (13) a. $c: VERB = \{\pi: i, \Sigma:be.here\}$ b. $c: VERB = \{\pi: li, \Sigma:be.there\}$
- (14) a. c: AUXILIARY = { π : í, Σ :?} b. c: AUXILIARY = { π : *l*í, Σ :?}

The verb–auxiliary multifunctionality is a case where a lexical category does double duty as a grammatical category. But patterns of multifunctionality are not restricted to this type. We also find cases where a single form may instantiate two different types of grammatical categories. It is, for example, a common pattern across unrelated languages that demonstratives serve double duty as complementizers. This is illustrated on the basis of English in (15); representations of the two instances of this UoL are given in (16).³

- (15) a. I know that guy.b. I know that this guy is courageous.
- (16) a. *c*:DEMONSTRATIVE = { π :*that*, Σ :?} b. *c*:COMPLEMENTIZER = { π :*that*, Σ :?}

The patterns of multifunctionality illustrated here are often viewed as a result of a grammaticalization path (Heine 1994; Heine and Kuteva 2002; Hopper and Traugott 2003; see Roberts and Roussou [2003] and van Gelderen [2004] for a generative approach towards grammaticalization). But the grammaticalization approach is not itself an explanation for the affinity between certain categories or why certain UoLs are more prone to a recategorization than others. Moreover, the fact that similar grammaticalization paths are attested across unrelated languages suggests that there is something universal about these *recategorization* processes. And consequently, we may conclude that there is something universal about *categorization* processes.

The postulation of a categorial label that mediates between a UoL and its interpretation serves as a necessary step towards an explanation for the pervasive patterns of multifunctionality. The pervasiveness goes beyond these patterns of polysemy we have just observed. Other patterns of multifunctionality that are determined by the syntactic context include *expletives* (loss of interpretation), *syncretism* (one UoL occupying multiple cells within a paradigmatic organization), and *fake forms* (partial loss of interpretation). We shall see instances of these patterns of multifunctionality throughout this monograph. What they share in common is that the same UoL is interpreted one way in one syntactic context but another way in a different syntactic context. Since the syntactic distribution of a particular UoL is an indication

³ From a descriptive point of view, the demonstrative version of *that* seems to have more semantic content than the complementizer *that*: the former includes a notion of displacement (*there* rather than *here*) which is absent in the latter. Leu (2008) argues that this displacement feature in demonstratives is supplied by a silent *there* which is overtly realized in many languages (see Section 6.4.3.2. for discussion).



Figure 1.2 Categorial identity mediates the relation between a UoL and its interpretation

of its categorial identity we may conclude that c affects the interpretation of a given UoL, as illustrated in Figure 1.2.

If this is on the right track, we have in turn evidence for the linguistic reality of c.⁴ In the course of exploring what this reality looks like we will address the question as to how exactly it influences the interpretation of a given UoL and in so doing we will be able to shed some light on the nature of grammatical meaning.

Crucially, patterns of multifunctionality of this sort can be used as universal diagnostics for categorical patterns.

1.1.2 Patterns of contrast

A second diagnostic for the presence of a category that can be universally applied has to do with the classic structuralist notion of *contrast* (Trubetzkoy 1939). As Saussure famously argued, language is defined by contrast: '*Dans la langue il n'y a que des différences*... *sans termes positifs*' ['In a language there are only differences, and no positive terms'] (Saussure 1967 [1916]: 166).

To see contrast at work, consider English plural marking in (17). While the plural is marked with the suffix -s, the singular is morphologically unmarked. And crucially, this unmarked form is not compatible with a plural interpretation as evidenced by its incompatibility with a numeral of cardinality greater than 1, as shown in (18).⁵

- (17) a. They planted the bug.b. They planted the bug-s.
- (18) a. *They planted three bug.b. They planted three bug-s.

⁴ Whether this linguistic reality corresponds to a psychological reality as well is a different question that I will set aside here. See Cohen and Lefebvre (2005) for relevant discussion on this issue.

⁵ Though as we shall see immediately below, the unmarked form is not universally associated with a singular interpretation. Rather unmarked forms are often interpreted as an instance of *general number*.

Table 1.1 A paradigmatic contrast

Base	Marked by	Interpreted as
N	$\{\pi: \emptyset, \Sigma: singular\}$	singular
Ν	$\{\pi:-s, \Sigma: plural\}$	plural

Table 1.2 An interpretive contrast

Base	Marked by	Interpreted as
N	-	singular
Ν	$\{\pi:-s, \Sigma: plural\}$	plural

So how does the unmarked noun trigger a singular interpretation? A number of answers have been proposed. They can be classified into two types. On one view, the singular interpretation arises in the presence of a dedicated UoL, which enters into a *paradigmatic contrast* with the overt plural marker but happens to be zero. This is illustrated in Table 1.1.

On the other view, the singular interpretation arises solely due to the absence of plural marking. This is illustrated in Table 1.2.

There are several ways to derive the presence of what appears to be a dedicated interpretation in the absence of a dedicated UoL. The singular interpretation can be considered a *default* that need not be directly encoded (Harley and Ritter 2002). Or else it may come about as an instance of Gricean-style reasoning (Sauerland 2008). This is grounded in the assumption that speakers are always as specific as possible. Thus, in light of the absence of plural marking, an addressee may conclude that the speaker must intend a non-plural interpretation, and non-plural equals singular. Essentially the same idea can also be modeled as a morphological principle instead of a pragmatic one, namely in terms of *the blocking principle*.⁶ Only the most specified form compatible with a particular interpretation can be used. So even though the unmarked form may in principle be compatible with a plural interpretation, the existence of a more specified form (the plural marked form) blocks its use.

⁶ This is also known as the elsewhere principle, Panini's principle, or the subset principle (Kiparsky 1973; DiSciullo and Williams 1987; Noyer 1992; Williams 1994, 1997; Halle 1997; Wiese 1999; Stump 2001).

Table 1.3 Two ways of being unmarked

Base	Marked by	Interpreted as	Markedness status
N	_	general number	truly unmarked
N	{ <i>π</i> :Ø, Σ:singular}	singular	zero marked
N	{ <i>π</i> :-s, Σ:plural}	plural	overtly marked

No matter how the interpretive contrast in Table 1.2 is implemented, it faces a fundamental problem. Not all morphologically unmarked forms are interpreted as singular. For example, in the context of a compound (19), the unmarked noun (bug) is not associated with a dedicated singular interpretation. Instead it is compatible with a plural interpretation. A bug spray is not a spray against a single bug as is obvious from the continuing sentence in (19).

(19) **Bug spray** won't help. There are bugs everywhere.

This suggests that not all unmarked nouns trigger a singular interpretation. The compatibility of the unmarked form with both a singular and a plural interpretation is sometimes referred to as *general number* (Corbett 2000; Rullmann and You 2006). The accounts based on interpretive contrasts may still be rescued, however. Since plural marking is not allowed within compounds, the unmarked form is not blocked in this context. However, unmarked nouns in Halkomelem Salish are compatible with a plural interpretation (20a), even outside of compounds where plural marking can otherwise occur (Wiltschko 2008).

(20) a. te lhíxw swíweles DET three boy 'the three boys'
b. te lhíxw swóweles DET three boy.PL 'the three boys'

Wiltschko 2008: 642 (3)

This indicates that nouns not marked as plural are not all treated equally: some unambiguously trigger a singular interpretation, while others are compatible with both a singular and a plural interpretation. We thus have to recognize two ways of being unmarked (Table 1.3). In Wiltschko (2008), I argue that unmarked nouns in Halkomelem (and inside English compounds) are truly unmarked (i.e., they are not marked as singular) while singular nouns in English are marked as singular, albeit not by an overt UoL, but instead by a zero marker (\emptyset).



Figure 1.3 Categorial identity mediates between form and interpretation

But how do we distinguish between a zero-marked noun and a truly unmarked noun? We are facing yet another problem of multifunctionality: at least on the surface, the UoL used for the singular interpretation is identical to the UoL used for the general number interpretation. And again, this difference is syntactically conditioned. In the context of a compound an unmarked noun is truly unmarked while in the context of a nominal phrase (introduced by a determiner) an unmarked noun is in fact marked as singular. As with other cases of syntactically conditioned multifunctionality, we can model this difference by postulating the presence of a category. In particular, the singular interpretation in the absence of overt marking is indicative of the presence of a category, while the absence of a dedicated interpretation (general number) is indicative of the absence of such a category. Furthermore, associating the singular interpretation with a categorial identity predicts that there are other instances of the same category. And this is indeed the case: plural marking is another instance of the same category, which is typically identified as c:NUMBER. This is schematized in Figure 1.3.

The presence of a categorial identity not only mediates between form and interpretation and licenses zero marking; it also is syntactically active in that it participates in syntactic relations (such as agreement).

This contrasts with the Halkomelem pattern where unmarked forms are always truly unmarked and therefore are never associated with categorial identity. This is consistent with the fact that the plural marker does not form a class with other UoLs, zero marking is not available, and it does not participate in syntactic relations (Wiltschko 2008). Instead the plural marker, a simple sound-meaning correspondence (marked as $\{\pi, \Sigma\}$) in Figure 1.4 combines with a noun to trigger the plural interpretation.

We have now seen two types of patterns where the presence of a categorial identity mediates the relation between UoLs and their interpretation:

- (i) multifunctionality
- (ii) zero marking licensed by a categorial contrast.





Figure 1.4 Direct mapping between a UoL and interpretation

In both cases, the UoL acquires a distinct interpretation by being associated with a particular categorial identity (c). These patterns are attested across categories and across languages, as we shall see. This suggests that UoLs are categorizable as a matter of the universal language faculty, i.e., *Universal Grammar* (UG). But at the same time, c does not seem to be an intrinsic property of UoLs in that they can exist without a categorial identity. This lies at the heart of both categorical patterns. Multifunctional UoLs are intrinsically without c but may be classified by two (or more) different categorial identities. And truly unmarked UoLs instantiated by general number nouns may or may not be classified as c:NUMBER. The question is how does c come about?

In what follows I discuss two opposing answers. On the one hand, we have the *Universal Base Hypothesis* according to which *c* comes about as a matter of UG. Accordingly, UG makes available a set of universal categories. This hypothesis, however, faces problems in light of much variation in the categorial inventories of the languages of the world. This is the starting point for the opposing view, which I dub the *No Base Hypothesis*, according to which there is no set of universal categories. I discuss each of these hypotheses in turn.

1.2 The Universal Base Hypothesis

In this section I introduce the Universal Base Hypothesis and the problems it presents.

1.2.1 The universal base as a repository of categories

The Universal Base Hypothesis (henceforth UBH) goes back to the early days of generative grammar (Chomsky 1965; Bach 1968; Lakoff 1970; Ross 1970 [1968]), but has been revived and updated over the years. In its early days, the *base* comprised both lexical rules and phrase structure rules. The former set of rules was responsible for the *categorization* of words (e.g., N \rightarrow *dog*) whereas the latter was responsible for *word order* (e.g., NP \rightarrow Det A N). According to early instantiations of the UBH the base is identical across all languages, with