The economy of word classes in Hiw, Vanuatu

Grammatically flexible, lexically rigid

Alexandre François

CNRS – LACITO, Paris / A.N.U., Canberra

The issue of lexical flexibility is best tackled as the articulation of two separate mappings: one that assigns lexical items to word classes; another one that associates these word classes with the syntactic functions they can access. A language may endow its lexemes with more or less multicategoriality, and its word classes with more or less multifunctionality: these are two distinct facets of lexical flexibility, which should be assessed separately. Focusing on Hiw, an Oceanic language of northern Vanuatu, I show that lexical flexibility is there mostly due to the high multifunctionality of its word classes, each of which can regularly access a broad array of syntactic functions. Conversely, Hiw ranks relatively low on the scale of multicategoriality: most of its lexemes are assigned just one word class. This is how a language can be grammatically flexible, yet lexically rigid.

1. On lexical flexibility

The notion of LEXICAL FLEXIBILITY measures the ability, for individual lexemes in a language, to fill a number of different syntactic functions in the sentence.[1] A language will be assigned a higher degree of flexibility if it allows a larger number of functions to its lexemes. By contrast, the other end of the typological spectrum will include languages that can be described as more "rigid", as they only allow their lexemes to fill one specific function.

A language like Latin shows low flexibility: words are normally assigned a single category (noun, adjective, verb…) and a root can only change category membership by means of morphological derivation. Thus, from the root *tim-* 'fear', Latin would derive

a verb (*timeō* 'to fear'), an adjective (*timidus* 'fearful') or a noun (*timor* 'fear'); each of these lexemes, taken individually, is assigned a very limited set of syntactic functions.

By contrast, Oceanic languages are often seen as more flexible in their lexicon. For example, Tahitian (Polynesian) has a word *'ite* which can function as a transitive verb 'see, know'; as an adjective meaning 'knowledgeable'; or as a noun meaning 'knowledge' or 'witness' (Vernaudon, n.d.):[2]

(1) a. ⟨'Ua ***'ite***⟩ Hina i te tai'o.
    CPLT know H. OBL ART read
    'Hina KNOWS how to read.'
    [*'ite* = head of TAM-inflecting predicate ≈*verb*]

   b. ⟨E feiā ***'ite***⟩ rātou.
    PRED people know 3PL
    'These are KNOWLEDGEABLE people.'
    [*'ite* = modifier of head in nominal predicate ≈*adjective*]

   c. ⟨E mea rahi⟩ tōna ***'ite***.
    PRED thing big POSS:3SG know
    'Her KNOWLEDGE is immense.'
    [*'ite* = head of argument phrase ≈*noun*]

The contrast between word classes in Tahitian is less straightforward than in European languages (Lazard & Peltzer 2000: 23; Vernaudon & Rigo 2004). Many categories are found in similar syntactic contexts: nouns, adjectives and verbs can be the head of a predicate marked for Tense-Aspect-Mood (1a); all three can be preceded by a determiner (1c), and so on. A language like Tahitian would rank high on a typological scale of lexical flexibility.

The question central to this volume is whether the lexical flexibility observed for a language like Tahitian can be generalised to the whole Oceanic family, or if Oceanic languages show a lot of variation on this scale. The present paper will look closely at Hiw, a non-Polynesian language of the Oceanic family spoken in Vanuatu [§2.4]. I will show that the grammar of Hiw shows both signs of high flexibility and high rigidity; and will attempt to solve this paradox.

During the course of this case study, I will also touch upon matters of theory and methodology that may be useful to future research on other languages, or on the issue of cross-linguistic comparison. Our discussion will try to answer three questions in particular:

---

2. Throughout the examples of this study, I will occasionally indicate the limits of the predicate phrase with pointy brackets ⟨…⟩.

a. What exactly is lexical flexibility? does it happen in the lexicon, or in the grammar?
b. How can lexical flexibility be assessed empirically, within the system of one language?
c. How can lexical flexibility be compared across languages?

2. Walk on two legs: Lexical mapping, grammatical mapping

2.1 From lexemes to word classes, from word classes to functions

In the first lines of this paper, I proposed a preliminary definition of LEXICAL FLEXIBILITY as "the ability, for individual *lexemes* in a language, to fill different syntactic *functions*". In fact, I would argue that it is problematic to define this notion as though it simply involved a direct mapping from lexemes to functions. I believe it is both heuristic and more accurate to introduce here an intermediate step, namely the notion of *parts of speech* or *word classes*.[3]

What might have appeared, as a first approximation, as a single mapping – from lexemes to functions – is in fact a two-step process:

a. *Lexical mapping*: LEXEME → WORD CLASS
    Each individual lexeme is assigned one or several word classes.
b. *Grammatical mapping*: WORD CLASS → FUNCTIONS
    Each individual word class is assigned one or several functions.

Thus in English, consider the word *paper*. While it typically heads an argument phrase as in (2a), it can also function as a modifier to another head, as in (2b):

(2) a. I need *paper*.
     b. I need a *paper* towel.

Should we say that *paper* shows lexical flexibility? that it behaves sometimes as a noun, and sometimes as an adjective? Hardly. In English, it is a property of all common nouns to be compatible both with the function of head in an argument phrase, and with that of modifier.[4] Thus *paper* can take on two different syntactic

---

3. My reflection on these matters owes considerably to the work of Alain Lemaréchal, especially his 1989 volume on parts of speech, *Les parties du discours* (Lemaréchal 1989).

4. In English, phrases consisting of two nouns routinely lexicalise as compounds, whether written in one or two words: *nightclub*, *school fees*, *time bomb*… (Ryder 1994). But whichever their degree of lexicalisation, the underlying syntactic structure of these compounds (cf. Benveniste 1974 [1967]) is always ⟨N$_{modif}$ N$_{head}$⟩, in which the second noun is the head, and the preceding one is a modifier (Bauer 1998).

functions, not because it is a flexible lexeme, but merely by virtue of being a noun. The fact that nouns can access not just one but two functions in English is worthy of notice; but rather than constituting a case of lexical flexibility as such, it would be more accurate to describe it as an example of GRAMMATICAL FLEXIBILITY, as it were. It is a property of word classes in the grammar, not a property of individual words in the lexicon.

This is not to say that English lacks lexical flexibility altogether. Consider the word *ship*, which can occupy the functions {head of argument phrase}, {modifier of head in argument phrase}, and {head of TAM-inflected predicate}:

(3) a. This *ship* is unsinkable.
     b. I collect *ship* models.
     c. They will *ship* your books by plane.

This example illustrates both lexical and grammatical flexibility. On the one hand, (3a) and (3b) reflect the multifunctionality of the word class Noun in English,[5] just like we saw for (2a)–(2b) above: this pertains to grammar, and says little about the lexicon. Yet on the other hand, (3c) features the same word as {head of TAM-inflected predicate} – a function that English does not normally associate with Nouns, but with Verbs. The best analysis is to consider that the lexeme *ship* maps onto two different word classes: it is a Noun – which accounts for (3a) and (3b) – and it is a Verb – which explains (3c). This double membership does constitute, this time, a proper case of lexical flexibility (assuming we take this expression literally, as referring to the lexicon).

A second example of lexical flexibility in English would be a word like *home*. It behaves like a Noun in (4a) and (4b), but like an Adverb in (4c), normally the only word class that can directly fill the syntactic slot of adjunct:

(4) a. My *home* is yours.
     b. This is my *home* country.
     c. There's nobody *home*.

Once again, what we have here is a dual lexical mapping (*home* → {NOUN; ADVERB}), followed by a grammatical mapping that can be either simple (ADVERB → adjunct) or dual (NOUN → {head in argument phrase; modifier in argument phrase}). The English situation is summarised in *Figure 1*.

---

5. Throughout this paper, I will capitalise the name of parts of speech. This usage is based on a structuralist notion, that the "nouns" in one language are in principle a different sort of reality from the "nouns" in another language. This point has been argued, among others, by Haspelmath (2010, 2012), and is in fact central to this whole study [§6.1]. The typography intends to reflect this analysis (cf. Haspelmath 2010).
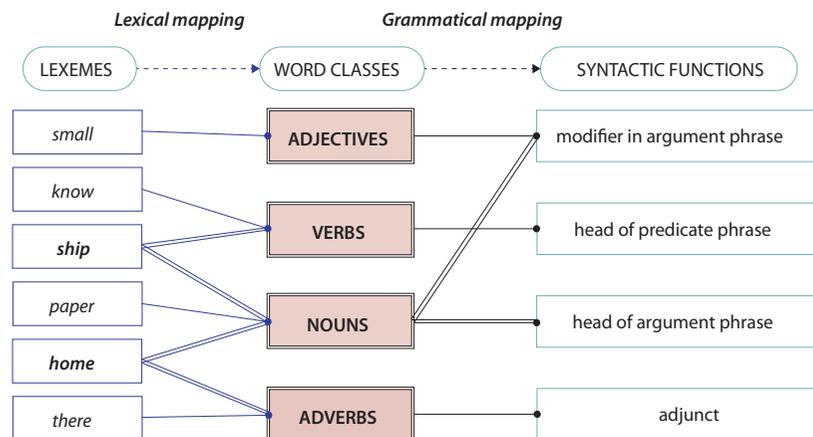
Figure 1. The two-step mapping from words to functions: English

2.2   Lexical vs. grammatical flexibility

The contrast between grammar and lexicon can also be framed in terms of predictability vs. arbitrariness. Within the grammatical system of English, it is entirely predictable that a given word, once it has been identified as a Noun, will be able to fill the two functions {head of argument phrase} as in (3a), and {modifier of head in argument phrase} as in (3b). This dual syntactic behaviour does not need to be memorised in the lexicon, because it is not specific to a particular lexeme: it is a general property of the word class Noun in the grammar.

By contrast, the fact that the same form *ship* can also head a TAM-inflected predicate in (3c) is not a predictable property that could be derived from its nominal status: while *ship* can fill that function, many nouns (e.g. *paper*) cannot. The difference in behaviour between the two words *ship* and *paper* cannot be predicted based on their form or their meaning, nor from any grammatical property: it is an arbitrary characteristic of these individual words, which therefore must be stored in the lexicon. The lexical properties of the word *paper* include its word-class membership as {NOUN}, whereas the form *ship* has dual membership {NOUN; VERB}.

The precise semantic relationship between the noun *ship* and the verb *ship* also shows some degree of arbitrariness, because it cannot be predicted systematically or derived regularly, neither based on semantics or on structure; again, it must be learnt and memorised in the lexicon. For that reason, it is sensible to consider the two pairs <*ship*, NOUN> and <*ship*, VERB> as two separate linguistic entities. Then there can be debate on how to call these entities. One may consider we're dealing

with a single lexeme (or lexical entry) /*ship*/, itself split into two separate "sublexemes" each endowed with a different part of speech, and with its own meanings. While the various senses of a single sublexeme would constitute *polysemy*, the mutual relationship between two sublexemes is best described as *heterosemy*[6] – when "different but related meanings of a given morpheme are associated with distinct grammatical contexts" (Enfield 2006: 197). Alternatively, one may prefer to describe <*ship*, NOUN> and <*ship*, VERB> as two different lexemes, in a mutual relationship of *conversion*. Both analyses may be valid, depending on one's definition of *lexeme*. But the crucial point here is that the assignment of more than one word classes to a single word form, and the precise semantic relation between the resulting linguistic signs, constitute non-predictable pieces of information that must be stored in the lexicon.

Ultimately, what we call *flexibility* can therefore refer to two very different phenomena.

Within the lexicon, flexibility corresponds to the arbitrary assignment of word-class membership to individual lexical items, with more or less predictability in the semantic relation between the different (sub)lexemes involved. This phenomenon is also known as *heterosemy* or *conversion*. The semantic contrast between *ship* in (3a) 'a nautical craft' and in (3c) 'transport s.th. to a remote location' pertains to the lexicon.

Conversely, grammar is essentially the realm of regularity and predictability. If the meaning of a certain form adapts to its syntactic context in a predictable way, then the semantic calculation takes place in the grammar, and does not involve storing or retrieving information in the lexicon. For example, the semantic shift between *ship* in (3a) 'a nautical craft' and in (3b) '[model] *of* a nautical craft, resembling a n. c.' can be entirely inferred from the grammatical construction $\langle N_{modif}\ N_{head}\rangle$. It is a matter of grammar that, in this language, lexical nouns can be referential (such as when they head an argument phrase) but also can describe a non-referential, intensional property (such as when they modify another noun in an argument phrase).

2.3   Multicategoriality vs. multifunctionality

This two-mapping view will allow us to disentangle the intricacy of lexical flexibility. I propose to break up the gradient of flexibility into two different metrics, which I will call *multicategoriality* and *multifunctionality*:

---

6. For the notion of *heterosemy*, see Lichtenberk (1991), Enfield (2006). The term *sublexeme* is my coinage, and is not used by these authors.

(5) the scale of MULTICATEGORIALITY reflects the ability, for individual lexemes, to be assigned several word classes

(6) the scale of MULTIFUNCTIONALITY reflects the ability, for individual word classes, to be assigned several syntactic functions

Multicategoriality is a property of the lexicon; *multifunctionality* a property of the grammar. By extension, both metrics can be seen as properties of a language as a whole.

A language like Latin would be a case of a maximally rigid system, as it shows low ratings on both dimensions. Most lexemes are assigned a single part of speech (e.g. *timeō* is rigidly a verb), and most parts of speech are restricted to essentially one or two functions (e.g. finite verbs can only be a predicate head).

At the other end of the flexibility scale, a language like Tahitian – briefly mentioned in §1 – combines multicategoriality [§6.3.4] with high levels of multifunctionality [§6.3.2].

English fits somewhere between these two extremes. As Figure 1 shows, it appears quite rigid for many lexemes (*small* is strictly an adjective, *know* strictly a verb, etc.), yet it does show some degree of multicategoriality (cf. *ship*, *home*).[7] English also shows a certain degree of multifunctionality – illustrated here for its nouns – but this remains quite limited: unlike Tahitian for example, English does not allow its nouns or its adjectives to head a predicate, or its verbs to head an argument.

It would be worthwhile to design a quantitative method for actually rating languages along these two metrics. While the detail of such methods falls beyond the scope of this study, I will make some proposals in the final discussion [§6.3].

2.4 Assessing lexical flexibility in Hiw

The present article will focus on the description of one language, Hiw (see François 2010a, 2012). This is an endangered language spoken by about 280 people on the island of the same name, at the northwestern tip of the Vanuatu archipelago, in the Torres Islands [Figure 2]. This small island group is also home to Lo-Toga, a language very close to Hiw (François 2010b, 2014a: 182).

---

7. Note that the term *precategoriality* sometimes used in the literature would be inadequate. If a word like *ship* were analysed as "precategorial", we would still have to explain why it never behaves like an adjective or an adverb. It is therefore more accurate to speak of *multicategoriality*, whereby a given lexeme maps onto specific word classes: *ship* is both a verb and a noun, *home* both a noun and an adverb. As for the term *precategorial*, it is probably better suited for roots, at a sublexical level (cf. Verhaar 1984, Lehmann 2008) – see in §1 the example of the Latin root *tim-*.
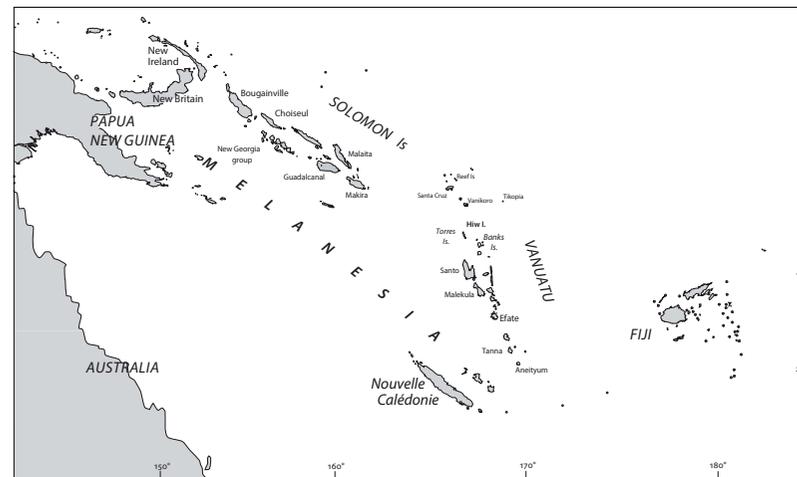
Figure 2. The island of Hiw (Torres Is.), at the northern tip of the Vanuatu archipelago

The present study rests on primary data collected by the author during several field trips to the Torres Islands, between 2004 and 2011. My corpus includes conversations and narratives (17,600 words), as well as elicited material.[8]

I will attempt to assess the forms taken by lexical flexibility in Hiw. Following the reasoning in the previous sections, it is clear that this type of study will first require us to establish the inventory of word classes in this language. Because such an inventory is always language-specific, the analysis will have to be conducted on a structuralist, distributional basis [§6.1]: I will examine the behaviour of each word class in my corpus, and list the array of syntactic functions they can regularly access in the clause.

Observing the general behaviour of parts of speech in the language's grammar – that is, what I called patterns of *grammatical mapping* from WORD CLASSES to FUNCTIONS – is a prerequisite before we can survey individual lexical items, and assess the patterns of *lexical mapping* from LEXEMES to WORD CLASSES. (Likewise, we needed to know about the syntactic properties of nouns in English, before being able to assess the lexical flexibility of individual words like *paper* or *ship*.)

---

8. I will indicate the source of my examples using simple conventions. Sentences taken from my recorded texts will note the story and the sentence number – e.g. [Meravtit.051]. Sentences obtained through elicitation refer to my field questionnaires – e.g. [d12:12]. Spontaneous speech heard during language immersion has a reference to my notebooks – e.g. [FP3-28b]. (My field notes are archived online, at *www.odsas.net*.)

For example, consider the pair of sentences in (7a–b). It shows that the same form *mařenage* can act as the head of an argument phrase (7a), but also as the head of a predicate phrase inflecting for tense (7b).

(7) a.   MAŘENAGE   mmo   ti   tuwtōw.
         chief   sick   PAST   before
         'The *chief* was sick before.'

     b. Noke   MAŘENAGE   ti   tuwtōw.
         1SG   (be).chief PAST   before
         'I *was a chief* before.'   [GP1:77]

A superficial, anglocentric analysis could be tempted to describe these two functions, respectively, as "nominal" and "verbal"; and even infer that *mařenage* is a noun 'chief' in (7a), and a verb 'be chief' in (7b). This analysis would be legitimate, for example, if it could be shown that nouns are normally incompatible with the function of (TAM-inflected) predicate head: we would then have a case similar to (3c) for English, and good reasons to analyse it as a dual mapping (*mařenage* →{noun; verb}). This would be an instance of lexical conversion, or multicategoriality.

But the analysis becomes entirely different if it turns out that {head of TAM predicate} is in fact a syntactic function open to all nouns. This would mean that this ability to fill that function is not encoded among the special characteristics of the lexical item *mařenage*, but is simply a general property of the class of Nouns in Hiw. The semantic shift between the use as head of an argument phrase ('the [one who is] N') and the use as head of a predicate phrase ('be N') is entirely derivable from the meaning of the respective syntactic functions (cf. §2.2). This particular meaning does not need to be encoded anywhere in the lexicon: there is here no reason to speak of conversion or heterosemy. In other words, we are dealing with multifunctionality, a feature of the grammatical system rather than the lexicon.

In this paper, I will propose a similar analysis not just for this Example (7a–b), but for many syntactic configurations. What may seem, at first glance, to constitute lexical flexibility in Hiw, is often an optical illusion (if we take *lexical* literally, as meaning "taking place in the lexicon"). While it is true that lexical items are able to occupy many syntactic slots in the sentence, this is generally not due to flexibility inside the lexicon, but rather to the very wide array of functions that are made accessible to each part of speech in the grammar. In other words, what characterises Hiw is first and foremost a high level of grammatical multifunctionality.

By contrast, and somewhat surprisingly perhaps, this language ranks relatively low on the scale of multicategoriality. Just like the word *mařenage* is strictly a noun, likewise the majority of lexemes in Hiw are assigned just one word class. Multicategoriality (conversion) does exist – for example, we will see in §5.1 that

veřoye is both a noun 'war' and a verb 'to wage war' – but it applies only to a couple dozen words, and is much more limited than, say, in English. In sum, the present study will show that the Hiw language is *lexically rigid* (low multicategoriality) yet *grammatically flexible* (high multifunctionality).

2.5   This study

The present article is organised as follows.

Section 3 will survey the syntax of the clause in Hiw, and present the formal clues for identifying the main syntactic functions. Section 4 will then define the different word classes of Hiw, based on their compatibility with these functions; this will provide an assessment of the language's multifunctionality. Section 5 will then survey the lexicon in search of multicategorial words. While these do exist, I will show that their number is in fact quite limited.

Finally, the last section will discuss how the method used here for Hiw can be generalised to other languages; and how we can compare languages with respect to lexical flexibility.

3.   The main syntactic functions of Hiw

3.1   The clause in Hiw

Before we examine the properties of individual parts of speech in Hiw, it is useful to begin with an overview of its basic syntax. This section will examine how the clause is organised in this language, and what are the formal correlates – e.g. in terms of word order – of its main syntactic constituents.

Like most other Oceanic languages, Hiw is a configurational, right-branching language in which word order tends to be strict. Hiw follows a nominative-accusative syntax: the subject (S) of monovalent predicates always aligns formally with the agent-like participant (A) of bivalent predicates: S/A precedes the verb, whereas the patient-like participant (O) follows it; S and A share a single set of pronominal forms distinct from O; etc.

The basic constituent order of Hiw would be traditionally described as SVO. However, in this study I will deliberately refrain from labelling basic constituents using terms inherited from word class labels: thus rather than parsing the sentence into the classical units of formal syntax "NP" and "VP", I will choose to employ functional labels such as "argument phrase" or "predicate phrase". This is a cautionary step in order to avoid any aprioristic bias regarding the nature of each phrase's

head (cf. Hengeveld 1992; Hengeveld & Van Lier 2010). For example, in sentence (7b) above [§2.4], rather than describe the constituent *mařenage ti* as a Verb Phrase, I will use the neutral label "TAM-inflected predicate": this has the advantage of leaving open all options as to the nature of its head (noun, verb or otherwise). And indeed, my conclusion will be that /*mařenage ti*/ 'was a chief' is headed by a Noun [§4.7.7]: describing it as a "Verb Phrase" would be inaccurate and ethnocentric.

The order of the main components in a Hiw clause comes as follows:

(8)   (TOPIC) (SUBJECT)$_{ArgP}$ PREDICATE$_{PredP}$ (ADJUNCTS)

The PREDICATE PHRASE (PredP) may in turn parse into a head and its modifiers:

(9)   PredP → { HEAD (MODIFIERS) }$_{PredP}$

If the predicate head is transitive, then it takes an object argument, which inserts inside the Predicate phrase:

(10)   PredP → { HEAD (MODIFIERS) [OBJECT]$_{ArgP}$ }$_{PredP}$

Hiw does not have double-object constructions. Any complement other than the direct object is treated as an adjunct, usually through the use of prepositions.[9]

The language has Differential object marking, treating human objects differently from non-humans [see Example (30)]. This structure has some morphological intricacies (François 2014b), which won't be detailed here.

Both the subject and the object take the form of an ARGUMENT PHRASE (ArgP). The latter may also parse into a head followed by its optional modifiers:

(11)   ArgP → { HEAD (MODIFIERS) }$_{ArgP}$

This study will focus on five main syntactic functions:

– head of predicate phrase
– modifier in predicate phrase
– head of argument phrase
– modifier in argument phrase
– adjunct

The following subsections [§3.2–3.6] will illustrate these syntactic functions, and provide more detail about their formal manifestations. Little will be said, quite deliberately, about the word classes compatible with these functions, as this is the topic of Section 4 below.

---

9.   See Example (87) below for the equivalent of a ditransitive construction; and Example (28), (29), (45), (73), (77), (92) for other prepositional adjuncts.

3.2   Head of a predicate phrase

A well-formed clause consists of at least one obligatory phrase, the PREDICATE:

(12)   ⟨MEŘAWE⟩.[10]
        perfect
        'It's perfect.'

Hiw has regular *zero* anaphora for non-human subjects – as in (12). All other cases require an overt subject. This takes the form of an argument phrase [§3.4], which always precedes the predicate:

(13)   Söřö   ⟨PUN⟩.
        3DU   quarrel
        'They argued.'   [d08.Rao:00]

The predicate can carry overt marking for Tense-Aspect-Mood. I will call these cases, for the sake of brevity, "TAM predicates":

(14)   Sise   ⟨**në**   KËKKË⟩.
        3PL   STAT   small~INTSF
        'They're very small.'   [Devils.02]

TAM markers in Hiw can precede the head, like Stative *në* in (14). They can follow the head, like Past *ti* in (7b) above, or in (15):

(15)   Kemi   ⟨MOTŘIG   **ti**⟩   vo?
        2PL   sleep:PL   PAST   where
        'Where did you guys sleep?'   [Meravtit.135]

Finally, certain TAM markers take the form of a discontiguous morpheme ⟨TAM$_1$… TAM$_2$⟩, like the Background Perfect[11] *ve… ti* in (16):

(16)   Tuñwuyegë   ⟨**ve**   ŘAK   **ti**⟩.
        HUM:FEM:PL   BKPF$_1$ make   BKPF$_2$
        'The women made it.'   [Stories.106]

The two components of such bipartite TAM morphemes form a bracket around a string of elements that come in a strict order, including the predicate head and its modifiers [§3.3].

---

10.   Whenever relevant, I will capitalise the syntactic head of certain phrases (whether in the predicate or in an argument phrase).

11.   On the semantics of the Background Perfect in Hiw, see François (2010b).

3.3 Modifier in a predicate phrase

3.3.1 *Internal syntax of the predicate phrase*

The head of a predicate phrase (PredP) can take one or several modifiers internal to the phrase. The order of constituents within the PredP was foreshadowed in (10) above; (17) shows it in more detail:

(17) ⟨ (TAM₁) **HEAD** (*modifiers*) [object]ₐᵣ₉ₚ (TAM₂) (directional) ⟩ₚᵣₑ𝒹ₚ

The modifiers within a predicate phrase always follow their head:

(18) Ike ⟨tati TITTÖM **tnēg** ⟩   pëne.
       2SG NEG think   too.much about.it
       'Don't worry **too much** about it.'   [d04.41]

If the predicate phrase is headed by a transitive verb, the modifiers normally insert between the verb and its object argument:

(19) Ike ⟨R̄YË   **wetewate** [ne yö   ēñwe]ₐᵣ₉ₚ⟩!
       2SG sweep   clean   ART inside house
       '(You) sweep the room **clean**!'   [EP3 13b]

The head of a predicate phrase may be followed by more than one modifier:

(20) Tite   ⟨YË **tōuř wuřog**⟩ tom tite ⟨PYË   **teřog tgō**⟩
       1inc:PL see follow properly COMP 1inc:pl attach.bait trying firm

       tom ne pyë mik   ñwot.
       COMP ART bait APPREH break

       'Let's *make sure* we *try to attach* the bait firmly so it doesn't come off.'
                                                              [d11.Wora:19]

This example will be explained in §4.5 below.

3.3.2 *Contrasting adverbs and adjuncts*

Care must be taken to distinguish between those modifiers that are internal to the PredP – like *tnēg* in (18) – and those that fall outside of it – like *pëne* in (18). The literature often uses the term "adverb" to describe similar words, a label which is notoriously vague and ill-defined. Hiw draws a clearcut distinction here between two distinct parts of speech, which I will call respectively "adverbs" and "adjuncts".

The distributional contrast between the two slots is made evident when the PredP includes more material. For example, (21) includes a postverbal TAM₂ marker *ti*, followed by a spatial directional *vēn* 'up'. The latter two – as per (17) above – mark the right boundary of the Predicate phrase:

(21) Meřavtit ⟨VËN **wřog** ti   vēn⟩ *eřëne*.
       M.   ascend through PAST up   above
       'Megravtit managed to climb *through* to the top.'   [Meravtit.209]

A sentence like (21) makes it clear that *wřog* is internal to the PredP, whereas *eřëne* is external to it. Each of these lexical items is strictly bound to its own syntactic slot: *wřog* 'through' can never appear outside the PredP, and *eřëne* 'above' never inside. I propose to keep the term ADVERB [see §4.5] for those lexemes that are only found within the boundaries of the PredP – e.g. *wřog* 'through' or *tnēg* 'too much'. Conversely, those words which, like *pëne* in (18) or *eřëne* in (21), can only occupy an adjunct slot outside the PredP, will be called ADJUNCTS. In Hiw, lexical adverbs and lexical adjuncts constitute two watertight categories. Adjuncts typically express location in time or space [§3.6, 4.6], whereas adverbs encode various other meanings, particularly manner.

As we'll see in §4.5, the syntactic function { modifier of head in a PredP } can be filled by a variety of lexical categories: either lexical Adverbs – which are specialised in this function – or other parts of speech, especially Verbs and Adjectives. To take the example of (20), we'll see that the modifier *tōuř* is lexically a verb serialised to the head *yë* 'see', and that *tgō* 'firm' is lexically an adjective; whereas *wuřog* and *teřog* are lexical adverbs, specialised in this modifying position. Yet regardless of their lexical nature, these different word classes are all to be analysed here with the same syntactic function, namely, { modifier of head in a PredP }.

3.4 Head of an argument phrase

Argument phrases are used as core arguments of a verb (subject, object), as the object of a preposition, or as a possessor.

An argument phrase (ArgP) can be headed by a personal pronoun – cf. *söřö* '3DU' in (13), *sise* '3PL' in (14), *kemi* '2PL' in (15)… – or by a proper noun – like *Meřavtit* in (21). The case of nouns is more intricate.

Nouns split into two noun classes, depending on their syntactic behaviour in argument phrases. STRONG NOUNS are able to form the head of an ArgP, like *mařenage* in (7a), or *temařëřë* in (22):

(22) { TEMAŘËŘË }ₐᵣ₉ₚ ve   kay   me.
       old.man   IPFV crawl hither
       'There's an old man coming here.'   [Meravtit.121]

By contrast, WEAK NOUNS are unable to form directly the head of an ArgP: in order to do so, they require a determiner of some sort, like the article *ne*. While *temařëřë* 'old man' was a Strong noun, its synonym *tamesō* is a Weak noun:

(23) a. *{ TAMESŌ}ᵣₑfₚ ve   kay   me.
old.person IPFV crawl hither
'There's an old man coming here.'

b. { *ne* TAMESŌ }ᵣₑfₚ ve   kay   me.
ART old.person   IPFV crawl hither
'There's an old man coming here.'   [Meravtit.129]

I will come back to the distinction between Strong and Weak nouns in §4.7. For the time being, suffice it to say that only Strong nouns can form directly the head of an argument phrase.

3.5   Modifier in an argument phrase

A fourth key function in the clause is that of { modifier of the head in an ArgP } – for example, a noun's attribute. As shown in (11), this slot follows the lexical head of the ArgP:

(24) ne   wake *kkë*
ART boat little
'a small boat'   [Meravtit.45]

This modifier function can be filled by Adjectives, Weak nouns and Numerals.

3.6   Adjuncts

The adjunct phrase normally comes after the predicate phrase, as in (18) and (21) above. Adjuncts can also be topicalised:

(25) *Köñ   së*,   sise vën   tom   se   r̄ekove *yöte*.
day   INDF 3PL go:PL COMP 3PL work   in.garden
'*One day*, they went to work in the garden.'   [Eel.64]

The adjunct phrase can consist of a prepositional phrase, a phrasal adjunct, or a lexical adjunct. In (25), *köñ së* is a phrasal adjunct; *yöte* is a lexical adjunct [see §4.6, 5.2].

4.   Defining the word classes of Hiw

4.1   Grammatical flexibility in Hiw

The five functions we examined in the previous section were defined using functional and syntactic criteria, independently of the word classes that can populate them. As stated in §3.1, this was a conscious choice to avoid making any aprioristic assumptions about the distribution of word classes in the language.

These key functions provide us with a grid that will help us, precisely, define the parts of speech of the Hiw language based on an empirical, distributional approach. Section 4 will define each word class in this language by the specific array of syntactic functions it can access. *Table 1* shows the correspondence between Hiw's word classes and the syntac-tic functions they can regularly occupy.

Table 1. Major word classes in Hiw and their syntactic functions

Syntactic function	Verb	Adjective	Numeral	Strong noun	Weak noun	Adverb	Adjunct
head of argument phrase	–	–	+	+	–	–	–
modifier in argument phrase	–	+	+	–	+	–	–
head of TAM-inflected predicate	+	+	+	+	+	–	–
head of direct predicate	+	+	+	+	–	–	–
modifier in predicate phrase	+	+	–	–	–	+	–
adjunct	–	–	–	–	–	–	+

This table displays what I called earlier [§2.1] the *grammatical mapping*, from word classes to functions. What is striking is the relatively high number of functions that can be occupied by a given word class directly, i.e. with no need of derivation or extra morphology. In other words, Hiw shows a high degree of syntactic multifunctionality. I will come back to this table in the final discussion [§6.3].

The next subsections will examine word classes one after the other, based on my corpus, so as to provide evidence for the claims made in Table 1.

4.2 Verbs

A large subset of the lexicon (over 25% of lexemes) falls under the word class Verb. These verbs may be transitive or intransitive, have static or dynamic semantics.

4.2.1 *Head of a predicate phrase*

The principal function open to verbs is { *head of a predicate phrase* }. This normally entails the presence of a TAM marker that precedes or follows it [see §3.3]:[12]

(26) Ne tamesō ñot ⟨ve **TU**⟩.
ART old.person INDF IPFV stand:NPL
'There was an old man standing.'   [Eel.07]

(27) Nine ⟨**MITIR̄** ti⟩ yöñwe.
3SG sleep:NPL PAST at.home
'She slept at home.'

Hiw also allows its verbs to head a predicate phrase with no TAM marking:

(28) Sör̄ö ⟨**TU**⟩ yö tapego tuwē.
3DU stand:NPL LOC mat one
'They're standing on the same mat.'   [Hades.45]

A structure like (28) may receive two interpretations. On the one hand, it could be analysed as a case of a TAM-marked predicate in which the TAM happens to be *zero*. On the other hand, it may also qualify as what I will later call a "direct predicate". Indeed, we'll see that such a category (direct predicate, uninflected for TAM) needs to be posited to account for equational predicates in the domain of nouns [§4.7.6]. Given the existence of this type of predicates in the system, it may be wise to consider that sentences like (28) constitute a direct predicate. This reasoning explains why Table 1 assigned verbs to two slightly different functions: {head of TAM-inflected predicate}, and {head of direct predicate}.

---

12. Note, in passing, that Hiw encodes the category of verbal number, or pluractionality, in the radical of certain verbs (François 2009). The choice of the radical varies with the number of the verb's absolute argument (S/O). Thus, 'sleep' is (27) *mitir̄* for a singular or dual subject, but (15) *motr̄ig* for a plural subject; 'go' is (55) *tō* for SG/dual, (79) *vën* for plural. This is indicated in the gloss, respectively, as "sleep:NPL" (non-plural) vs. "sleep:PL". This suppletive pattern concerns thirty lexical pairs, a high number by typological standards (Veselinova 2013). Only verbs are concerned by this alternation in Hiw – unlike its neighbour Lo-Toga, which has this pattern both for verbs and for a few adjectives. I won't detail this point further here, as it bears limited impact upon our main discussion.

4.2.2 *Modifier in a predicate phrase*

Another function that can be filled by Hiw lexical verbs is {modifier of the head in a predicate phrase}. This happens when the verb comes second in a serial construction. Thus the verb *tu* 'stand' can be serialised to another verb *r̄ōw* 'dash, fly, leap', yielding a compound *r̄ōw tu* ⟨dash – stand⟩: [of an arrow] 'fly and reach a certain spot, get stuck somewhere':

(29) Ne mesor̄ =ena ⟨R̄ŌW **tu**⟩ i ne r̄öt pake.
ART arrow =his dash stand:NPL OBL ART root banyan
'His arrow *stuck* in the banian root.'   [Meravtit.199]

Likewise, (30) shows how *moketog* 'release' can be serialised to *mañe* 'talk', resulting in a compound ⟨talk – release⟩ 'authorise s.o. to do s.th.':

(30) Sise ⟨MAÑE **moketog** i noke⟩.
3PL talk release DOM 1SG
'They've authorised me.'   [EP3 04b]

Hiw has several syntactic structures which can be analysed as verb serialisation.[13] The one illustrated in (29)–(30) consists in stringing together two verb lexemes so as to form a "macro-verb", a compound verb made up of several phonological words.[14]

Just like many languages with serial verb constructions (Durie 1997: 322; Aikhenvald 2006a: 30, 45; François 2004: 134–136; 2006: 227), there is a continuum from patterns where a verb is productively used as a modifier to another verb, to fully lexicalised V V compounds such as the examples just given. But in all cases, each verbal component remains a distinct phonological word, and retains its morphological and syntactic properties (valency, affixal morphology: cf. François 2014b). Each compound mentioned here is thus phrasal rather than a single word. And just like N N compounds in English have their internal syntax ⟨N$_{modif}$ N$_{head}$⟩ (see note 4), likewise the lexicalised serial patterns of Hiw are still to be analysed, structurally, as a syntactic sequence ⟨V$_{head}$ V$_{modif}$⟩.

In sum, while this structure can be analysed as verb serialisation, it also follows a more general syntactic template in the language, that of a head followed by its modifier [§3.3]. The analysis sometimes proposed for verb serialisation as a multi-headed construction (e.g. Baker 1989) does not seem to fit the case of Hiw. The first verb V$_1$ always has the privileged status typical of a head: it is the only

---

13. François (2010b: 511–512) briefly presents these serial structures, which are shared by Hiw and Lo-Toga. François (2004, 2006) is a more detailed account of serial verb constructions in the neighbouring language Mwotlap. See also Crowley (2002) for other languages of Vanuatu.

14. This pattern would be described as *nuclear-layer serialisation* in the framework of Foley & Olson (1985) or Crowley (2002).

obligatory item in the structure, and always assigns the primary case frame of the whole phrase, particularly its syntactic subject. Serialised verbs are capable of affecting the valency of the head, but then this is true of all types of adverbial modifiers, and is not unique to verbs in a serial construction.[15] As far as semantics are concerned, the ⟨V$_1$ V$_2$⟩ phrase always refers to a subtype of V$_1$: e.g. *mañe moketog* 'authorise' is a particular case of 'talk'. In those structures, V$_1$ has the same array of meanings as when it is used on its own. This contrasts with the V$_2$ position, which for many lexical verbs entails a semantic shift: for example, *tōur̃* means 'follow (s.o., s.th.)' when used as the head of a predicate, but '[do s.th.] carefully' when used as a V$_2$ in a serial pattern.[16] In sum, serial verb constructions in Hiw are 'asymmetrical' (Aikhenvald 1999; 2006a: 21) constructions, best analysed as a structure consisting of a head (the main verb V$_1$) followed by a modifier (a verb V$_2$).

The syntactic slot {modifier in a predicate phrase} is also open to Adjectives, Numerals or Adverbs [see §4.5]. In fact, many Adverbs can be shown to originate in lexical verbs that have become specialised in this slot (cf. Aikhenvald 2006b: 197; François 2004: 137, 2006: 229): this common change path is another argument for considering that the V$_2$ slot in a serial pattern is just one particular instance of a more general structural slot of *adverbial modifier* – a slot open to a whole range of word classes, and not just verbs.

As far as the grammatical mapping [WORD CLASSES → SYNTACTIC FUNCTIONS] is concerned, we can therefore conclude that the function {modifier in a predicate phrase} is open to verbs.

4.2.3   *Other functions*

Verbs in Hiw are not allowed to modify directly the head of an argument phrase:

(31) **ne   tayö   **tu**
      ART   person   stand:NPL
      **the standing person*

The only way for a verb to modify an argument head within an ArgP would be in a relative clause, using the relativiser *pe* and a TAM-inflected predicate:

(32) ne   tayö   [pe   v'   ag   **tu**   r̃ë]$_{RelC}$
      ART   person   REL   IPFV   inland   stand:NPL   DEM
      'the person who's standing over there inland'

---

15. See François (2000, 2004: 116–140) for the argumentation regarding Mwotlap, whose serial patterns are parallel to those of Hiw.

16. The same is true of adjectives [§4.3] used in the same syntactic slot: e.g. *kkë* used as a predicate head means '(be) small', but when serialised to a main verb, it acts as a diminutive or attenuative [see Example (35)].

Likewise, verbs in their bare form cannot head an argument phrase. In order to refer to the event itself, a Hiw verb cannot simply combine with the article *ne* (as in *ne tu 's.o.'s standing'): it must be formally nominalised [§5.1.2], either through a suffix *-ove* (→ *ne tu-*OVE *nome* 'your standing') or through morphological reduplication.

As far as our survey is concerned, the category Verb is thus compatible with only three syntactic functions: {head of direct predicate}, {head of TAM-inflected predicate}, and {modifier in a predicate phrase}.

4.3   Adjectives

I have collected over a hundred lexical adjectives in Hiw, covering various semantic domains. These include: *yuy* 'white' and all colour terms; *at* 'alive', *mët* 'dead'; *sa* 'bad'; *wye* 'good', *vöwye* 'true'; *kkë* 'small', *pwö* 'big', *mesō* 'ripe, large'; *tr̃öt* 'tasty'; *tar̃otr̃ot* 'poor'; *yog* 'married'; *toqe* 'pregnant'; *mmo* 'sick, painful'; *r̃aqe* 'new'; *taëtwë* 'identical'; *tënnēr̃* 'adequate'; *ter̃oñye* 'easy', *tër̃ënta* 'difficult'; *vogmamer̃ö* 'sad'; *vr̃iwane* 'funny'.

Just like Verbs, Adjectives can head a TAM-inflected predicate. This is illustrated by the Stative aspect *në* in (14) above, or the Complete aspect *piti* in (33):

(33) Nine ⟨**pwö** piti⟩.
      3SG   big   CPLT
      '[the eel] it's already large.'   [Eel.25]

An adjective can also head a direct predicate, with no TAM marking:

(34) Vë~n vën,   mer̃ëmpē   ⟨ **pwö** ⟩.
      CONT CONT   eel   big
      'Over time, the eel got big.'   [Eel.15]

Also like verbs, adjectives can modify the head of a predicate phrase. In (35), the adjective *kkë* 'small' modifies the head *mesō* 'large':

(35) Ne   got   ⟨peon   MESŌ   **kkë**⟩.
      ART   food.parcel   FUT   large   small
      'The portions should be *slightly* larger.'   [Eel.41]

The semantic function of adjectival modifiers is sometimes to qualify the head – like the attenuative use of *kkë* 'small' in (35) – and sometimes to indicate a result:

(36) Nine pō   vën   vën   nine ⟨PŌ   **mët**⟩.
      3SG   slender CONT   CONT   3SG   slender dead
      'He lost weight, lost so much weight that he died.' [GG1-05a]

An adjective can modify a head which is itself another adjective – as in (35) and (36) – or it can modify a verb head – as in (19) above, repeated here:

(19)   Ike   ⟨R̄YË   *wetewate*   ne   yö   ēñwe⟩!
       2SG   sweep   clean   ART   inside   house
       '(You) sweep the room *clean*!'   [EP3 13b]

Unlike verbs (cf. 31), Adjectives can modify the head of an argument phrase:

(37)   TEKÑWA   *meyigeyige*
       HUM:MX:PL   black
       'black people'

(38)   ne   ñwë   *wye*   mi   ne   ñwë   *sa*
       ART   demon   good   with   ART   demon   bad
       'the good demon and the bad demon'   [Yams.25]

However, adjectives cannot themselves head an argument phrase, even with the help of the determiner *ne*:

(39)   **kkë*   /   **ne   kkë*
       small   ART   small
       *a/the small one…

For an argument phrase to be built around an adjective rather than a noun, it will require a dummy head of the form *në* similar to English 'the one' (etymologically from *ne gë* 'the thing'):

(40)   NË   *kkë*
       the.one   small
       'a/the small one'   [Meravtit.190]

In sum, adjectives in Hiw can fill four functions: {head of direct predicate}; {head of TAM phrase}; {modifier of head in predicate phrase}; {modifier of head in argument phrase}.

It may be tempting to describe Hiw adjectives as a subclass of verbs, based on their ability to form a predicate, whether direct or marked in TAM (33, 34). The faculty of adjectives to modify a predicate head (35, 36) would simply be a case of verb serialisation, which would be unproblematic. Simply, one would need to specify that this subclass of verbs can modify a noun (38), contrary to other verbs.

The reason I don't propose this analysis is that the properties shared by adjectives and verbs are also shared by numerals and, in part, by nouns (Table 1, §4.1) – particularly, the ability to form a TAM predicate. If adjectives are to be considered a subclass of verbs on the basis of the properties they share, then the same should

be said of numerals or nouns – to the point where virtually all word classes would end up being defined as a subclass of verbs, with the addition of extra (non-verbal) properties. Little would be gained with such an analysis. It thus seems more useful to consider verbs and adjectives as simply two distinct word classes – albeit ones that share a couple of properties. The commonalities shared by several parts of speech can equally be captured by acknowledging that Hiw, like most Austronesian languages, is "omnipredicative" – i.e. has a grammar in which most word classes can constitute the head of a predicate (Launey 1994; Lemaréchal 1994: 153; Evans & Osada 2005: 359).

4.4   Numerals

Numerals in Hiw form a closed lexical class which has its own set of possible functions. They commonly occupy the slot {modifier in an argument phrase}, like adjectives:

(41)   ne   köñ   *vīr̄ö*
       ART   night   two
       'two nights'   [Devils.15]

Yet contrary to adjectives, numerals can directly head an argument phrase:

(42)   *VIR̄Ö*   ⟨yër̄e   pe⟩.
       two   be.absent   now
       'Two are missing now.'   [Meravtit.180]

They can form direct predicates, with no need of a copula:

(43)   Ne   megoye   =na   ⟨*VIR̄Ö*   ñwutuye⟩.
       ART   child   =her   two   just
       'She has only two children.'
       [lit. 'Her children *are two only*.']   [Meravtit.156]

Numerals can even head a PredP that inflects for TAM – e.g. with the Subjunctive:

(44)   Ike   go   ne   wnot   ⟨on   *VIR̄Ö*⟩.
       2SG   wrap   ART   parcel   SBJV   two
       'You should wrap two parcels.'
       [lit. 'You wrap parcels so they *will be two*.']   [Eel.40]

My corpus doesn't have examples of a numeral modifying the head of a PredP. One does occasionally find a numeral in that position, but it is then reduplicated with a distributive meaning (*tuwë* 'one' → *tuwtuwë* 'one by one').

(45) Sise peon köge ***tuwtuwë***   i   ne   ti-r̄ēgye.
       3PL FUT tie   one~DISTR OBL ART leaf-Cordyline
       'They will tie them up *one by one*, using Cordyline leaves.'   [d10.Bekem:09]

Should we then say that Numerals can modify a predicate head? This would be far-fetched, considering the morphological and semantic changes involved in such structures. It is safer to conclude that reduplication is here deriving a Numeral into something else, namely an Adverb.[17] As far as bare Numerals are concerned, they apparently cannot modify the head of a predicate.

Finally, another function that numerals cannot occupy is that of adjunct.

In sum, Numerals can fill the functions {head of argument phrase}, {modifier in argument phrase}, {head of direct predicate}, {head of TAM predicate}. It is the only word class with that exact profile (see Table 1, §4.1).

4.5   Adverbs

I call lexical Adverbs those words which can only function as a modifier inside the boundaries of a predicate phrase.[18] I have more than sixty different Adverbs in my corpus, including *wuyog* 'again', *ver̄og* 'also', *r̄ake* 'up', *sur̄* 'down'…[19] We've seen already some examples, like *tnēg* 'too much' in (18), *wr̄og* 'through' in (21), *ñwutuye* 'just' in (43). Adverbs contrast with Adjuncts, which appear outside the PredP [§3.3, 4.6].

We have seen that the function {modifier in a predicate phrase} can also be occupied by verbs [§4.2] or by adjectives [§4.3]. By contrast with these two parts of speech, adverbs are restricted to that particular function: unlike adjectives, they cannot appear inside an argument phrase; and unlike both verbs and adjectives, they are ill-formed to head a predicate phrase themselves:

(46) *Nine ⟨ *wr̄og* ⟩ …
      3SG   through
      *He went through… (?)

---

17. Morphological reduplication in Hiw commonly has the power to derive lexical items from one word class into another. For example, we'll see in §5.1.2 that it can nominalise a Verb.

18. In previous publications, the category I here call *adverb* has been named using the terms *adjunct* (Crowley 1982:162; François 2004, 2005a:139, 2006) or *postverb* (e.g. François 2011:216–222).

19. Like other languages in the Torres – Banks region (François 2015:147 sq.), Hiw distinguishes between a set of space directionals (e.g. *vēn* 'up', *uw* 'down') and words with a similar meaning that belong to the category of Adverbs (*r̄ake* 'up', *sur̄* 'down'). These two word classes differ in their syntactic distribution [§3.3.1].

These observations can be turned into a set of syntactic tests. If a lexeme is used as a modifier in a predicate phrase, it may be either a verb, an adjective or an adverb. Thus (20), reproduced here, has four such modifiers, whose word-class membership is in principle ambiguous:

(20) Tite   ⟨YË ***tōur̄   wur̄og***⟩ tom tite   ⟨PYË   ***ter̄og***
      1inc:pl see follow properly COMP 1inc:pl attach.bait trying

      tgō⟩ tom ne   pyë mik   ñwot.
      firm COMP ART bait APPREH break
      'Let's *make sure* we *try to fix* the bait *firmly* so it can't come off.' [d11.Wora:19]

However, eliciting these words in different contexts makes the following points clear:

– *tōur̄* is a verb, since it can head a predicate phrase, with the meaning 'follow'. In (20) it is serialised to another verb *yë* 'see', yielding the compound meaning ⟨see follow⟩ 'look carefully'. [This is structurally parallel to (30) above.]
– *tgō* is an adjective meaning 'hard, solid, firm', which as such can modify a noun. In (20) it is used adverbially, to indicate the manner of the main verb *pyë* 'attach bait'. [This is structurally parallel to (36) above.]
– *wur̄og* 'properly' cannot be used in other positions than this adverbial slot, so it is an adverb.
– *ter̄og* is also found only as a verb modifier, with a conative meaning '(do) tentatively, try'. It is therefore a lexical adverb too.

Some modern adverbs originate in former adjectives, or former verbs in a serial construction, which have ended up specialising in this adverbial function [§4.2.2].[20] Of course, as far as this study is concerned, the word-class status of a word is assessed purely based on its synchronical distribution in the modern language.

4.6   Adjuncts

The class of lexical Adjuncts consists of words whose main syntactic function is that of an adjunct in the clause. As explained in §3.3, these adjuncts are always external to the predicate phrase, and are never found inside its boundaries. Their default position is to the right of the PredP (after a TAM marker or a space directional if there is one); occasionally, adjuncts are topicalised.

---

20. See François (2006:225) for similar cases in Mwotlap.

The vast majority of lexical adjuncts are LOCATIVES.[21] These are lexemes referring to locations in space (e.g. *yönwe* 'at home', *yönwr̄ëwōn* 'in the bush', *ginëne* 'here'…) or in time (e.g. *tōwtōw* 'in the past', *qutukñwa-ëne* 'now', *mer̄ën* 'tomorrow'…):

(47) Ye   ⟨ve tō   ti   iy me⟩ **yönwe**?
      who:SG BKPF₁ go:NPL BKPF₂ in hither in.house
      'Who came in here, *in the house*?'   [Grouper.33]

(48) Ne vot   in   ⟨ve toge wate që me⟩ **qutukñwaëne**.
      ART stone ANAPH IPFV stay until still hither now
      'This rock still exists *today*.'   [Stories.009]

(49) **Tōwtōw**, ne   wane,   sise tat   wane   vitikeyë.
      formerly ART drink.kava 3PL NEG:IRR drink.kava casually
      '*In the old days*, the drinking of kava, that was not done casually.' [Stories.101]

This category also includes all place names:

(50) Sör̄ö peon yöy   *Vile*.
      3DU   FUT   stay:NPL Vila
      'They'll be staying *in Vila*.'

As we will see in §5.2, Hiw has a few words which pattern both as (Weak) nouns and as Locatives: e.g. *yöte* 'garden; in the garden', *wōnaye* 'road; on the road'. Yet that syntactic behaviour is not available to all locative words.

As mentioned in §3.6, adjuncts can also consist of a prepositional phrase, whether this preposition has a locative meaning (*yö* 'in, at'; *r̄ē* 'on'…) or a non-locative one (*mi* 'with', *pē* 'about', *ti* 'Dative', *i* 'Oblique'…).[22] When the object of these prepositions is already activated in discourse, it is indexed as a 3SG suffix on the preposition – e.g. *mi-e* 'with him/her/it'. Because the morphology of person suffixes in Hiw can be complex, the combination of a preposition with an anaphoric (3SG) suffix results in surface forms that are sometimes unpredictable, and need to be learnt by the speaker. It is safe to analyse these combinations, ultimately, as though they formed a monomorphemic word: *yö* 'in' → *yöne* 'in it, inside'; *r̄ē* 'on' → *(e)r̄ene* 'on it, above' (21); *pē* 'about' → *pëne* 'about it' (18); *ti* 'Dative' → *se* 'to him/her'. Insofar as these

---

21. This is so true, that a possible label for the category could have been Locative rather than Adjunct. The reason I am keeping the more abstract term, is because of a handful of words in that class whose meanings is not locative (in time or space), such as (18) *pëne* 'about it', or (76) *ie* 'of it'.

22. For examples of preposition phrases in Hiw, see sentences (28), (29), (45), (73), (77), (79), (87).

---

words can be considered unanalysable, they qualify for the status of lexical Adjuncts, in the same way as Locatives.

Lexical Adjuncts cannot head a predicate phrase:[23]

(51) *N̄wati-k   ⟨*Vile*⟩.
      brother-1SG   Vila
      *'My brother is *in Vila*.'

Likewise, they normally cannot modify directly a noun. In order to do so, they must be derived using a particle *te* (glossed ORIG for 'originative'):

(52) *tuñwuyegë   **Hiw** → tuñwuyegë *te*   **Hiw**
      HUM:FEM:PL   Hiw   HUM:FEM:PL ORIG Hiw
      'the women of Hiw'   [Stories.085]

(53) *ne   wiywiy **tōwtōw** → ne   wiywiy *te*   **tōwtōw**
      ART habit   formerly   ART habit   ORIG formerly
      'customs of the past'   [Stories.105]

In sum, the only syntactic function that is productively accessible to all lexical Adjuncts is that of syntactic adjunct (see Table 1, §4.1).

4.7   Nouns

We now come to the domain of nouns, which is less straightforward than the other word classes of Hiw. Indeed, there are good reasons – already foreshadowed in §3.4 – for positing the existence of two noun classes, based on their behaviour in argument phrases.

4.7.1   *Strong nouns vs weak nouns*

It is expected that nouns in a language should be able to head an argument phrase. And indeed, this is the case for one class of nouns in Hiw. Take the example of *mar̄enage* 'chief' in (7a) above, *temar̄ēr̄ë* 'old man' in (22), or *ququy* 'friend' in (54):

(54) **QUQUY** =ena megoye penëne ve   yēr̄e.
      friend his   child DEM   IPFV be.absent
      'His child friend was not there.'   [Music.037]

---

23. This is one of the few differences between the system of Hiw and that of Mwotlap, which allows locative predicates (François 2003: 14, 2005: 128).

Proper nouns, incidentally, behave in the same way:

(55) *Sëkōp* tō piti.
Jacob go:NPL CPLT
'Jacob has already left.'

I propose to call "Strong nouns" those words which are – as it were – 'strong' enough to head an argument phrase on their own. Yet the majority of nouns in Hiw belong to the second class, which I will call "Weak nouns":[24] this class consists of nouns which are, strictly speaking, unable to form an argument phrase by themselves. Thus compare the examples of Strong nouns above, with the behaviour of *tamesō* 'old person' in (23) above, or of *megoye* in (56):

(56) a. **MEGOYE =na ⟨virö ñwutuye⟩.*
      child =her two just

   b. *NE megoye =na ⟨virö ñwutuye⟩.*
      ART child =her two just
      'She has only two children.'   [Meravtit.156]

In order to form a valid argument phrase, Weak nouns require to be preceded by a Determiner. This can be the article *ne*, or a preposed possessive classifier (57), or a Gender classifier (58) coding for gender and number [see §4.7.4]:

(57) Sörö r̃ak [NÖ-SA *megoye* tuwë].
   3DU make POSS-3PL child one
   'They gave birth to *a child*.' [lit. 'They made their child']   [Eel.02]

(58) [TÖR̃ÖQATE *megoye* nome] në wye ?
   HUM:MX:DU child your STAT good
   'Are *your two kids* alright?'   [Brothers.37]

4.7.2 *Syntactic analysis*
If argument phrases are to be equated with Noun phrases – as one could assume – how come they cannot be headed by nouns like *tamesō* or *megoye*?

   One way of analysing these structures is to equate the argument phrase not with a *Noun Phrase* (NP), but with a *Determiner Phrase* (DP). Since Abney (1987), DPs have been understood – at least within the framework of Government & Binding theory – to form a syntactic entity distinct from NPs. While an NP is headed by a noun, a DP is headed by a Determiner (or other words inherently endowed with the ability to head a DP, such as Strong nouns). Under this analysis, the D head of

---

24. I thank Mark Donohue (pers. com.) for suggesting these terms to me.

---

the DP in a sentence like (56b) would be the determiner *ne*, itself modified by the noun *megoye*.

   The syntactic restrictions of Hiw are not unlike those found in some more familiar languages. In English, the head of a DP can be a pronoun (SHE *is asleep*), a proper noun (MIKE *is asleep*), or certain particular nouns, e.g. some kin terms (DAD/ GRANDPA *is asleep*). However, the majority of common nouns are ill-formed to constitute a DP by themselves, and require some sort of determiner to do so (**CAT is asleep* → THE *cat is asleep*). Likewise in Hiw, an ArgP (=DP) may be headed by a pronoun, a proper noun, or a noun like *ququy* in (54); but the majority of common nouns can only constitute an ArgP through the support of a determiner.

   I thus conclude that Hiw has not one word class "Noun", but two distinct classes – respectively, STRONG NOUNS vs. WEAK NOUNS. Their difference in grammatical behaviour is rendered in the form of two separate columns in Table 1 (§4.1), where the function {head of argument phrase} is only associated with Strong nouns.

4.7.3 *Semantic definition of the two classes*
The distribution of nouns across the two nominal classes is not entirely random, and follows some semantic tendencies. The Strong nouns of Hiw are all semantically human (e.g. *mar̃enage* 'chief', *ququy* 'friend', *aukë* 'uncle'…). The category of Weak nouns includes all other nouns: inanimates (e.g. *wake* 'boat'), animals (*powëge* 'shark'), but also some human nouns (*megoye* 'child', *yeqën* 'woman'…).

   Table 2 provides a comprehensive list of the Strong nouns of Hiw.[25] Table 3 gives an overview of the Weak nouns ('++' means that a given category includes many more lexemes).

   The split thus runs across the set of human referents: some are Strong nouns, others are Weak nouns. The assignment of a concept to one of these two nominal categories is sometimes arbitrary, as witnessed by the pair *temar̃ër̃ë* [STRONG] vs. *tamesō* [WEAK] for the exact same meaning 'old man', or the pair *keko* [STRONG] vs. *megoye* [WEAK] for 'child'. That said, the tendency is for Strong nouns to encode those human referents that rank higher on the individuation scale, especially kin terms like *mam* 'Dad'; these nouns really follow the patterns of proper nouns [see

---

25. By convention, a string of characters followed by a tilde refers to an obligatorily possessed noun, the radical of which is always followed by a personal suffix. Thus, the noun *mar̃u~* 'uncle…' requires a suffix (e.g. *mar̃u-k* 'my uncle'). As for its non-suffixable synonym *aukë*, it can be used alone as an address term (*aukë!* 'uncle!'); when possessed, rather than taking a suffix, it combines with a personal clitic (e.g. *aukë =kye* 'my uncle'). The assignment of nouns to the suffixable vs. non-suffixable categories is a matter of morphology; it is ultimately arbitrary, and stored in the lexicon. This distribution is orthogonal to the contrast between Strong and Weak nouns: suffixable nouns are found both in *Table 2* (*r̃ekña~* 'mother'…) and in *Table 3* (*mya~* 'hand'…). For this reason, it shall not concern us further in this study.

Table 2. The strong nouns of Hiw: a comprehensive list

Semantic type	Examples
most KINSHIP NOUNS	*ma, teta* 'Mum'; *mam, pepa* 'Dad'; *r̄ekña~* 'mother'; *tema~* 'father'; *keko* 'child, offspring'; *pepu, pup* 'grand-parent/grandchild'; *ñwati~* '[M] brother'; *takyē~* '[F] sister'; *tutut, tutva~* 'opp. sex sibling'; *manegō~* 'cross-cousin'; *aukë, mar̄u~* 'maternal uncle/nephew'; *weyuk* 'same sex sibling-in-law'; *nemas* 'opp. sex sibling-in law'; *qiyige* 'father-in-law/son-in-law'; *qoyga~* '[F] mother in-law'; *r̄oqogë* 'daughter-in-law, [M] mother-in-law'
some other human nouns	*ququy* 'friend'; *tuqunkë* 'children'; *mar̄enage* 'chief'; *temar̄ër̄ë* 'old man'; *maësë* 'old woman'; *Mema* 'Pastor'
GENDER CLASSIFIERS	*r̄ëtëgë* 'HUM:FEM:SG, woman'; *tekñwa* 'HUM:MIX:PL, people'… (§4.7.4)
pronouns, etc.	*ye* 'who:SG'; *itiye* 'who:PL'; *së* 'INDEF'; *në* 'dummy.N' (§4.3)
PROPER NAMES	*Tōr̄a* 'Andora'; *Sëkōp* 'Jacob' ++

Table 3. The weak nouns of Hiw: an overview

Semantic type	Examples
ALL inanimates	*r̄ër̄ë* 'tree, wood'; *vönyö* 'island, country'; *yöte* 'garden'; *ñwute* 'place'; *mya~* 'hand'; *ya~* 'name'; *to* 'year'; *mr̄ë* 'wrath'; *yöyñisñis* 'desire'; *vegevage* 'speech' ++
ALL animals	*powëge* 'shark'; *r̄ite* 'octopus'; *qr̄ē* 'dolphin'; *ñwate* 'snake'; *gusuwe* 'rat'; *tok* 'dog'; *sōgë* 'pig'; *gove* 'heron'; *sō* 'chicken' ++
MOST human Ns	*tayö / qin* 'person'; *tamesō* 'old person'; *teñwën* 'man, husband'; *yeqën* 'woman, wife'; *megoye* 'child, offspring'; *yumegov* 'young boy'; *ñweyeñwaye* 'young girl'; *tayö ywö* 'leader'; *vetvatego* 'teacher' ++
human-like Ns	*temët* 'ghost'; *ñwë* 'demon'; *Wu* 'spirit, God'

(55)], which are by definition individuated. By contrast, Weak nouns with human referents often correspond to qualitative properties, devoid of extension: for example, *yeqën* 'woman' or *megoye* 'child' refer to an intensional quality (the quality of being female, or of being a child) and are not necessarily individuated.

I propose that it is precisely the role of determiners to provide intensional notions with extension and individuation. Thus, the bare noun *yeqën* would represent the intensional notion 'womanly, female', whereas *ne yeqën* would refer to 'a/the [particular] woman'.

4.7.4   *Gender classifiers*

Hiw has a closed paradigm of words – arguably grammatical rather than lexical – whose syntactic behaviour espouses that of Strong nouns. This paradigm is a set of portmanteau morphemes which always refer to humans, and combine number

with gender. Among the three semantic features encoded by these morphemes (humanness, number, gender), the only one that is unique to this paradigm – and in fact, very seldom encoded in Oceanic languages – is that of gender. This explains my choice to call these morphemes *Gender classifiers*, as a short form for gender-and-number classifiers for human referents.

Specific to the paradigm are the following parameters:

– four numbers: singular, dual, (optional) paucal, plural
– three genders: masculine, feminine, mixed

Table 4 provides the full paradigm of gender classifiers in Hiw.

Table 4. The gender classifiers of Hiw

	Singular	Dual	Paucal	Plural
MASC	(ne qin)	tōr̄ate	tuwesate	teñwar̄e
FEM	r̄ëtëgë	tōr̄ör̄ë	tuwutgë	tuñwuyegë
MIXED	(ne qin)	tōr̄öqate	tuwesate	tekñwa

The Gender classifiers of Hiw are partly similar to the articles or pronouns of other languages, but are in fact a distinct sort of grammatical device; their closest equivalent in English would be certain hyperonymic nouns like 'person', 'people', 'women'. For example, the form *tuñwuyegë* in (52) above was glossed HUM:FEM:PL, meaning 'a set of human referents of feminine gender, with more than two members' – that is, '(the) women'. Earlier on in (37), the most frequent gender classifier, *tekñwa*, was glossed HUM:MX:PL – that is, a mixed group of several humans, generally rendered in English as 'people'. As for *tōr̄öqate* in (58), it was glossed HUM:MX:DU, meaning 'a pair of humans of mixed gender' – in other words, a married couple, or a brother – sister pair, etc.

Relevant to our main discussion is the fact that these gender classifiers have the same syntactic distribution as Strong nouns. They do not need any extra morphology to head an argument phrase (DP). They can be used absolutely (i.e. with no modifier):

(59)   **TUÑWUYEGË**   pyë   n'   ov;   alë   **TEÑWAR̄E**   var̄er̄age.
       HUM:FEM:PL   light   ART   fire   then   HUM:MASC:PL   climb
       'The women lit fires, while the men went to the gardens.'   [Eel.68]

Like most argument heads (except personal pronouns), Gender classifiers can have their own modifiers, e.g. adjectives [as in (37), (68)] or quantifiers:

(60) TUÑWUYEGË **trö** tati yöymerën wurog ne vevë.
    HUM:FEM:PL some NEG know   properly ART weaving
    '*Some women* don't know well how to weave.'   [Hades.49]

In all these cases, there is no ambiguity that the head of the argument phrase is the classifier. Crucially, Gender classifiers – which themselves function like Strong nouns – are commonly modified by a Weak noun:

(61) RËTËGË   **tamesō**
    HUM:FEM:SG old.person
    'the old woman'

Likewise in (58) above, *töröqate* was modified by *megoye* 'child', yielding the meaning 'two (mixed-gender) children':

(58′) TÖRÖQATE   **megoye**
    HUM:MX:DU child
    'two children'   [Brothers.37]

In such structures, the Gender classifier is again syntactically the phrasal head, followed by a nominal modifier. Yet by the same token, it can also legitimately be analysed as a determiner for the following noun – similar to an article that would encode gender and number. Both analyses are correct, and mutually compatible: the gender determiner is the head of the DP, which is modified by the noun that follows.

    A Weak noun with human reference encodes number by replacing the singular article *ne* with a non-singular gender classifier:

(62) **ne** ñweyeñwaye   → **tuñwuyegë** ñweyeñwaye
    ART girl   HUM:FEM:PL girl
    '(a/the) girl'   → '(the) girls'

(63) **ne** tamesō   → **tekñwa**   tamesō
    ART old.person   HUM:MX:PL old.person
    'an old person' → '(the) old people, (the) elderly'

4.7.5 *Modifiers in an argument phrase*
We saw in §4.7.1 that Strong nouns and Weak nouns differ in their ability to head an argument phrase. Interestingly, their distribution is reversed when it comes to the function {modifier in an argument phrase}.

4.7.5.1 *Weak nouns*
Weak nouns are able to act as modifiers in an argument phrase – a characteristic they share with Adjectives [§4.3]. They can modify a Strong noun – as in (54) above,

where the Weak noun *megoye* 'child' modified the Strong noun *ququy* 'friend'. They can also do so with Gender classifiers, which arguably form a subset of Strong nouns: we saw this in (58) with *megoye*, and in (61)–(63) with other nouns.

    In addition, a Weak noun can commonly modify another Weak noun – itself preceded by the article:

(64) ne   MEGOYE **teñwën** in
    ART child   man   ANAPH
    'the boy'   [Eel.05]

(65) ne   YEQËN **tamesō**   ñot
    ART woman old.person INDF
    'an old woman'   [Eel.74]

4.7.5.2 *Strong nouns*
My corpus doesn't feature Strong nouns in a modifying position. The only exception is when a Strong noun is preceded by a gender classifier used to encode number:

(66) ñwati-k   → **tekñwa**   ñwati-k   [Meravtit.076]
    brother-1SG   HUM:MX:PL brother-1SG
    'my brother' → 'my brothers'

In principle, the internal structure of a phrase *tekñwa ñwati-k* is one where *tekñwa* is the head, and *ñwati-k* the modifier; at least, this analysis was the most convincing one in the case of Weak nouns such as *tamesō* in (63). In this perspective, an example like (66) could be taken to show that Strong nouns can occupy the slot of modifier in an argument phrase, just like Weak nouns.

    However, it is worthy of notice that this construction ⟨Gender classifier + Strong noun⟩ is the only case where Strong nouns appear in second position in a phrase: other combinations – e.g. one where a Strong noun would modify another noun – are unattested. One may thus propose an alternate analysis for (66), namely that *ñwati-k* is the head of the argument phrase not only in the singular, but also when marked for number; under that analysis, the Gender classifier *tekñwa* in (66) would simply be there to indicate number, without affecting the internal structure of the argument phrase.

    In sum, there is legitimate doubt as to which term is the head in (66) *tekñwa ñwati-k*. For this reason, the evidence is not strong enough to conclude that Strong nouns can regularly function as modifiers in Hiw. I will conclude that they cannot – as represented in Table 1, §4.1.

4.7.6   Head of direct predicate

Like most Austronesian languages, Hiw lacks a copula: its nouns are directly pre-dicative. Yet we want to be more precise here: are both noun classes in Hiw equally predicative? Do they differ in their syntactic behaviour? They do.

Just like Strong nouns were able to head an argument phrase, they can also form the head of a predicate phrase, with no need of any morphology. Thus, a nominal predicate, whether it is equational (*X is [the] Y*) or ascriptive (*X is [a] Y*), follows in Hiw a simple pattern {X ⟨Y⟩} in which a subject DP (X) is followed immediately by the predicate (Y), which also has the syntactic structure of a DP:

(67)   Kamaře   ⟨*QUQUY*⟩.
        1exc:du   friend
        'We are friends.'   [Music.31]

(68)   Sise   ⟨*TEKŊWA   ain   maqu*⟩.
        3PL   HUM:MX:PL   different   completely
        'They are completely different people.'   [Stories.031]

By contrast, Weak nouns are unable to form a direct predicate – that is, a predicate with no TAM specification:

(69)   *Noke   ⟨*YEQËN*⟩.
        1SG   woman
        *I'm a woman.

In order to form a predicate, Weak nouns need to be preceded by a Determiner – typically, the article *ne*:

(69′)   Noke   ⟨**ne**   *YEQËN*⟩.
        1SG   ART   woman
        'I'm a woman.'   [d12.Sintia:16]

(70)   Ne   tamesō   in,   nine   ⟨**ne**   *TEMËTŘÕŊ*⟩.
        ART   old.person   ANAPH   3SG   ART   shaman
        'That old man, he was a shaman.'   [Yams.08]

The reader will be familiar with the pattern here. The only way to form an ascrip-tive predicate is by way of a DP, not an NP. By definition, Strong nouns are able to form DPs by themselves, so they can easily head a direct predicate too, as in (67). Conversely, Weak nouns are unable to form a DP unless they combine with a determiner; this rule, which applied in the case of argument phrases [§4.7.2], also applies for non-verbal predicates.

The same DP structure is required for two different types of nominal predicates. One corresponds to *ascriptive predicates* ('X is an N'), as shown in (69′) and (70) above. But the same structure is also required for *equative predicates* ('X is the N'):

(71)   Pa   ne   votwu   ne,   ⟨**ne**   VOTWU   =kye⟩.
        but   ART   knife   DEM   ART   knife   POSS:1SG
        'This knife is my knife.'

In sum, a strict analysis of Hiw entails that Strong nouns are predicative, but Weak nouns are not. This is not about the need of a copula – as Hiw does without any such device in all cases – but about the requirement for Weak nouns to combine with extra material in order to constitute a well-formed DP. In application of this strict analysis, the conclusion is that Weak nouns are unable to head a direct (non-TAM) predicate [see Table 1, §4.1].

4.7.7   Head of a TAM predicate

The situation is slightly different when the predicate is endowed with Tense-Aspect-Mood specifications – what I have called "TAM predicate" [§3.2]. This is in fact the only syntactic function which is open equally to both categories of nouns, whether strong or weak.

4.7.7.1   Strong nouns

Strong nouns can form a predicate by themselves, as in (67) above. But to this facul-ty of being predicative, they also add the ability to combine with TAM specifications:

(72)   Tȫřö   ⟨**peon**   *QUQUY*⟩.
        1inc:du   FUT   friend
        'You and I will be friends.'   [Music.22]

From the semantic point of view, a nominal predicate inflecting for TAM is an ascriptive predication endowed with a time perspective. On the one hand, (67) above simply ascribed a nominal property to the subject (*we = friends*), yet said nothing about the time limits of that property. On the other hand, a TAM-inflected noun predicate like (72) assigns the property N to the subject at a certain point in time, contrasting it with other periods when *X was not yet N*, and/or *X will be no longer N*. Likewise, if the TAM marker is semantically modal, it sets a contrast between different possible worlds (a world where *X is N* vs. other worlds defined by *X is not N*, etc.):

(73)   Noke   ⟨**ta**   *QUQUY*⟩   mi-ke ?
        1SG   POT   friend   with-2SG
        'Can I be friends with you?'   [Hades.30]

While a direct (non-TAM) predicate will often translate in English as a present tense 'X is N', its TAM-inflected counterpart will usually involve a TAM-inflected copula in English ('X has been/was/will be/could be… N'), or a change-of-state predicate ('X has become N', 'X turned into an N'…).[26]

Due to these semantic restrictions, the TAM inflection of nouns tends to be mostly attested with certain nominal concepts which are inherently unstable – or rather, inherently compatible with an unstable reading. For example, being a 'house' is normally conceived as a stable property (something either is a house or is not), so this noun is not commonly found associated with TAM marking in Hiw. Conversely, that construction is more typically found with nouns depicting properties that do change over time – e.g. *ququy* '(be) friend', *mařenage* '(be) chief', *yumegov* '(be) young', *tamesō* '(be) old'…[27]

Crucially, even if only a small minority of lexical nouns are attested in my corpus as the head of a TAM predicate, this combination is always possible. ⟨TAM+ Noun⟩ is always grammatical, and easily interpreted, due to the transparent semantic relationship between the two structures. Knowing that a direct predicate means ⟨be an N⟩ or ⟨be the N⟩, a TAM predicate will simply be understood as a similar ascriptive or equative predication, but specified for a certain tense, aspect or modality. As a result, {head of a TAM predicate} can legitimately be seen as a function inherently open to all nouns, whether or not they are attested as such in my corpus.

Section §2.4 already cited an example of a TAM inflected noun in Example (7b), repeated here:

(7b) Noke ⟨MAŘENAGE   ti⟩   tuwtōw.
      1SG   chief   PAST before
      'I was a chief before.'

The question initially asked in §2.4 was whether (7b) was an instance of multicategoriality. It would be the case if *mařenage* were a member of a subclass of lexemes that would sometimes pattern as a noun, and sometimes as a verb, through a mechanism of conversion in the lexicon.[28] In fact, we now know that the function {*head of a TAM-inflected predicate*} is in principle – semantics permitting – open to all nouns in the language; and the semantic shift from 'chief' to 'be a chief' is entirely

---

26. See François (2003: 53–72; 2005b) for an in-depth analysis of TAM-inflected noun predicates in Mwotlap.

27. Among the four nouns cited here, the first two are Strong nouns, the latter two are Weak nouns (for which, see §4.7.7.2).

28. See the examples of multicategorial words given above for English (*ship*, *home*) and below for Hiw (§5).

and regularly predictable. There is no reason to consider that *mařenage* has been transformed into a verb in (7b): the form *mařenage* is as much a noun in (7b) as it was in (7a) when it was heading an argument phrase. In sum, a sentence like (7b) does not illustrate conversion, or lexical flexibility in the strict sense of the term, but *grammatical* flexibility – that is, the ability for a given word class (in this case, Strong nouns) to regularly fill a variety of syntactic functions.

This is parallel to the English example we discussed in §2.2: when an English noun is employed in a modifying structure, like *ship* in (3b) *ship model*, it doesn't stop being a noun. Its exact meaning in such a context can be entirely predicted from the general properties of the syntactic construction where it is used; there was no need to posit any conversion there, or zero-derivation, or heterosemy. Likewise, because Hiw is an "omni-predicative" language (cf. Launey 1994 for Classical Nahuatl), the predicative use of *mařenage* in (7b) remains perfectly "nominal" within the system of this particular language, with no good reason to speak of conversion or heterosemy. Given any noun lexeme N, it will always be possible for a speaker to calculate its meaning whether in an argument phrase ("…a/the N") or in a predicate phrase ("…is/was/will be a/the N"), with no need to resort to any adhoc "semantic increment" (cf. Evans & Osada 2005: 371).

In sum, the semantic flexibility that allows Hiw speakers to use and interpret nouns as heads of a TAM predicate is not to be located in the lexicon (properties of individual lexemes), but in the grammar (properties of word classes).

4.7.7.2   *Weak nouns*

We saw in §4.7.6 that Weak nouns cannot, strictly speaking, head a direct predicate – unless they are derived into a DP by means of some determiner.

However, this rule changes in the presence of TAM marking, which enables any Weak noun to form a predicate. For example, the Weak noun *yumegov* 'young boy' is ill-formed to head a predicate by itself, but it can do so when combined with the Stative *në* as in (74′):

(74) *Kimiře   ⟨yumegov⟩
      2DU   young.boy
      *You are *young boys*…

(74′) timeřën   pe   kimiře   ⟨në   YUMEGOV   që⟩
       time   REL 2DU   STAT young.boy still
       'at a time when you were both still *young (boys)*'   [d12.12]

Likewise, the Weak noun *köñ* 'night' commonly serves as a predicate to the subject *ñwute* (literally 'place'), to indicate the time of day, in which case it takes TAM inflection:

(75) Ne   ñwute   ⟨ᴋöñ   piti⟩.
     ᴀʀᴛ   place   night   ᴄᴘʟᴛ
     'Night has already fallen.'
     [lit. The place has already nighted.]   [d08.Rao:02]

And in a more figurative way, the same word *köñ* '(be) night' can take the word 'mind' as its subject (*yö~*), yielding a sentence that means 'to forget':[29]

(76) Ne   yö-k   ⟨ᴋöñ   piti⟩   ie.
     ᴀʀᴛ   mind-1sɢ   night   ᴄᴘʟᴛ   ᴀɴᴀᴘʜ
     'I have forgotten about it.'
     [lit. 'My mind has nighted about it.']   [d12.Sintia:13]

Formally speaking, it could be argued that the underlying head of a ᴛᴀᴍ predicate is really the ᴛᴀᴍ specification itself – a notion akin to the concept of *Infl* in Minimalist theory.[30] Under such an analysis, the actual "*Infl*" head of the predicate (74′) would be its Stative *në*, of which the noun *yumegov* would only be a specifier. Such a syntactic analysis would have the advantage of providing a consistent account of Weak nouns, which in all other constructions appear ultimately unable to head any major constituent: these nouns can only function as complements to a D head in a DP, whether it is an argument or a predicate. In the case of ᴛᴀᴍ predicates, it might well be that Weak nouns are really complements to the I (*Infl*) head in an IP. I will not go further into this reasoning, which is not essential to our discussion anyway.

4.7.8   *Summary: nouns*
The section on nouns was the most complex of our grammatical overview, and may warrant a brief summary.

Hiw does not have one but two categories of nouns – labelled here Strong vs. Weak nouns. These two categories share the same behaviour only with respect to the modifiers they can take on their right (adjectives, demonstratives, possessives, etc.). However, these two classes of nouns differ in their syntactic distribution, as they cannot occupy the same syntactic slots in the clause.

The only function that is shared by Strong and Weak nouns is that of head in a predicate inflecting for Tense-Aspect-Mood – see (72) and (74′). Functions exclusively filled by Strong nouns include: {head of argument phrase} and {head

---

29. This turn of phrase is common in northern Vanuatu, for several predicates relating to emotions and mental faculties (see François 2013: 204–209).

30. In Minimalist theory, a node called IP (with I for 'Inflection') has a category *Infl* ('inflection') as its head, and a VP as its complement (see Radford 1997: 65).

of direct predicate}. As for Weak nouns, they are normally restricted to modifier functions: they can modify a Strong noun, a Gender classifier (which itself behaves like a Strong noun), or another Weak noun. The most common construction in which they appear is in combination with the article *ne* (or other determiners), which is the device required to derive them into a well-formed DP – whether an argument phrase or a predicate.

Besides the main functions that have been discussed here, a number of other syntactic constructions treat Strong and Weak nouns differently. For example, prepositions form two distinct classes, depending on whether they take as their object a Weak noun ("light" prepositions) or a Strong noun ("heavy" prepositions). Likewise, the morphosyntax of possession differs widely depending on the nature of the nouns involved; etc. All these observations point to the same conclusion: Weak and Strong nouns constitute two distinct word classes in Hiw.

Finally, Hiw uses a set of Determiners (notably the article *ne*) as a productive strategy to turn a Weak noun, so to speak, into the structural equivalent of a Strong noun. Without going into the detail of each construction here, Table 5 (from François, in prep.) makes it clear that Determiners enable Weak nouns to fill all the syntactic functions that are normally open to Strong nouns. (Bolded rows in the table correspond to those syntactic functions that were discussed in this study.)

Table 5. Determiners enable Weak nouns to fill the syntactic functions of Strong nouns

Function	Weak nouns (bare form)	Weak nouns with prenominal Determiner	Strong nouns
	e.g. *megoye* 'child'	e.g. *ne megoye* (a/the) 'child'	e.g. *keko* (a/the) 'child'
object of associative modifier	+	–	–
incorporated object	+	–	–
object of light preposition	+	–	–
generic direct possessor	+	–	–
modifier of another N	+	–	–
ᴛᴀᴍ-inflected predicate	+	+	+
equational/ascriptive predicate	–	+	+
core argument of predicate	–	+	+
topic	–	+	+
appellative	–	+	+
object of heavy preposition	–	+	+
referential constructed possessor	–	+	+

To use the approach developped by Lemaréchal (1989) inspired by Tesnière (1959), one could say that the role of the Determiner is to productively *transfer* (Fr. "translater") a Weak noun into the class of Strong nouns. Syntactic *transfer* (Fr. "la translation") is an efficient strategy that allows the member of a given word class to easily access the functions normally licensed to a different word class.

Table 5 also highlights the usefulness of assigning categories not just to individual lexemes, but also to larger constituents. The grammatical behaviour of the class STRONG NOUN, at the level of lexemes, matches the behaviour of the phrasal category "Referential phrase" (DP), which can surface as {Det + Weak noun}. The analysis proposed in this paper for word classes could thus, in principle, be extended to phrases and constructions as a whole. However, I will not develop further this issue in the present study, which focuses on flexibility at the level of lexemes.

5. Multicategoriality in Hiw

In sum, what may seem to constitute, at first glance, lexical flexibility in Hiw, can in fact largely be explained by its high rates of multifunctionality, that is, GRAM-MATICAL FLEXIBILITY. Many syntactic constructions made it initially tempting to speak of conversion, or multicategoriality, and assign the mechanism of flexibility to individual items in the lexicon. Yet further scrutiny revealed that for many of those constructions, the lexeme's flexible behaviour was in fact a property of an entire word class, so that there was no ground at all for positing a change to a distinct class. For example, we saw nouns used in a context that could be assumed – based on European languages – to be "verbal" in nature (head of a TAM predicate); but given that those syntactic functions are productively accessible to any noun in the system, a better analysis is to simply consider that these predicative functions simply belong to the array of functions assigned to the word class Noun in this particular language.

Is this to say that Hiw lacks any LEXICAL FLEXIBILITY at all? This would be exaggerated. In addition to its high rates of multifunctionality, the language does also show genuine flexibility in the lexicon. This lexical flexibility is best defined as *multicategoriality* [§2.3] – that is, the ability for a single lexical form to belong to more than one word class. For reasons of space, I will not endeavour to provide a comprehensive discussion of multicategoriality in Hiw. In this section, I will simply illustrate a few multicategorial lexemes, so as to pursue my general methodological reasoning on assessing flexibility in a language.

5.1   Noun–Verb hybrids

At the beginning of this study, I illustrated multicategoriality using the example of English words like *ship*, which combine the properties of a common noun with those of a verb [§2.1]. Hiw also has a number of lexemes that can add up the grammatical behaviour of a Verb with that of a (Weak) noun. I will call them NOUN–VERB HYBRIDS.

5.1.1   *Presentation of noun–verb hybrids*
Table 6 provides a sample of such Noun–Verb hybrids.[31]

Table 6. Some noun–verb hybrids of Hiw

	Verb	*Noun*
* řekove*	'work'	'work'
řařōñ	'lament'	'lament'
genegone	'battle, wage war'	'battle, war'
veřoye	'fight'	'fight'
vegevage	'speak'	'speech, language'
togekëse	'play'	'game'
wane	'drink kava'	'kava drinking'
wëte	'rain'	'rain'
pyë	'attach bait'	'bait'

All these words illustrate conversion, or heterosemy [see §2.2]: a single form is assigned two distinct word classes – in this case, Verb and Noun. This dual mapping cannot be predicted: it is stored in the lexicon, as an arbitrary property of these particular lexemes.

Let's consider, for example, the lexeme *veřoye*. As a verb, it can form the head of a predicate phrase, with the meaning 'fight, carry a fight':

(77) Tekñwa   tuwtōw sise ⟨**veřoye**⟩ uř   ne   tën.
      HUM:MX:PL in.past 3PL fight   CAUS ART land
      [VERB] 'Our ancestors used to *fight* for land.'   [Stories.086]

The same form is also a noun. As such, *veřoye* can be preceded by the noun article *ne* – a property exclusive to Weak nouns – to form an argument phrase:

---

31. Being semantically inanimate, all nouns here are Weak nouns [see §4.7.3].

(78) Nine gat r̄ōw tom **ne** **ver̄oye** on pa p' ëne.
      3SG say out COMP ART fight SBJV end FOC DEM
      [NOUN] 'He ordered that *the fight* should stop right now.'   [Stories.075]

The same noun can be the object of the locative preposition *yö*:

(79) Sise vën ti **yö** **ver̄oye**.
      3PL go:PL PAST LOC fight
      [NOUN] 'They went into a *fight*.'   [Stories.004]

Similarly, *r̄ekove* can be used as a verbal predicate, meaning 'to work', as in Example (25), §3.6. When the same word functions as a Weak noun, it may form an argument phrase (in this case, the object of the verb):

(80) Nine take yar̄ **ne** **r̄ekove** r̄ōw Santo.
      3SG SEQ search ART work DIREC Santo
      [NOUN] 'Then she went to look for a *job* in Santo.'   [q.d03.Torba:15]

Another typically nominal function for *r̄ekove* is that of predicate:

(81) Ne r̄ak-vë-suqe   ⟨**ne** **r̄ekove** mesō⟩.
      ART make-NMLZR-initiations ART work big
      [NOUN] 'The preparation of initiations is a huge *work*.'   [Stories.048]

Because it is a Weak noun, *r̄ekove* here requires the article *ne* to form a predicate. (81) is a sequence of two DPs, a subject followed by its predicate [see §4.7.6]. Even though *r̄ekove* is here used predicatively, there is no question that it is here still a noun – as witnessed by the presence of a determiner and an adjective.

The syntactic contrast between (80) and (81) is a matter of grammatical flexibility, involving the various functions assigned to the word class Noun in the system; all (weak) nouns behave exactly the same in Hiw. Conversely, the contrast between *r̄ekove* in (81) and *r̄ekove* in (25) involves two separate word classes, respectively Noun and Verb. This is an instance of lexical flexibility strictly speaking, i.e. multicategoriality. Only a few nouns behave this way in Hiw, and this dual affiliation must be stored separately in the lexicon.

The same demonstration could be carried out for the other lexemes of *Table 6*. In those functions that are typically filled by verbs, they behave like any verb. Yet they are also compatible with the functions occupied by nouns in the language (modifier of Determiner in argument phrase; object of preposition…). What is special about these lexemes is not so much that they occupy several functions – after all, we've seen that multifunctionality is a staple of Hiw grammar. Rather, the main point here is that the array of functions attested for these lexemes does not

match any single word class, but combines the functions of two distinct classes – respectively that of verbs and of (weak) nouns.

As it happens, most of the nouns in Table 6 above can be understood as the *nomen actionis* for the corresponding verb: thus *r̄ekove* as a noun 'work' designates the activity defined by the verb *r̄ekove*. If this semantic shift were systematic, then we would be facing a case of grammatical flexibility, similar to the one we discussed in §4.7.7 above about nouns used as predicate heads, with no need to posit dual membership. However, the semantic properties of these Noun–Verb hybrids are not always predictable. For example, *pyë* [v.] 'fix bait' – [N.] 'bait' does not constitute a pair {verb; *nomen actionis*}; the noun refers not to the action of fixing a bait onto a fishing line, but to the bait itself.

Also, not all verbs can be converted into a noun: thus, while *ver̄oye* is a hybrid lexeme, *mañe* 'talk' is exclusively a verb, and cannot be used as a noun; likewise for *yō* 'see', *vën* 'go', *sag* 'sit', and most other verbs. Nothing in the grammatical system makes it predictable that a word like *ver̄oye* 'fight' could be used as a noun and a verb, while a verb like *mañe* 'talk' could not. The fact that a given form is a pure noun, a pure verb, or a Noun–Verb hybrid, needs to be learnt by the speaker: it is stored in the lexicon, as a property specific to certain lexical items. This is the sort of multicategoriality that defines, strictly speaking, the lexical component of flexibility.

5.1.2 *Hybrid lexemes vs. nominalising derivation*

While the words in Table 6 are only a sample of a larger list, it is necessary to point out that the number of Noun–Verb hybrids in Hiw is low. Out of a corpus-based list of 1188 lexemes, I was able to identify only 18 individual Noun–Verb hybrids: this is about 1.5% of the lexicon [§6.3.4], a very small rate compared to the amount of similar lexemes in a language like English. Verbs like *sō* 'fall', *toge* 'stay', *woge* 'cry', *tar̄e* 'cut', *tëp* 'love', *tō* 'walk'… are exclusively verbs, and are incompatible with any syntactic use as a noun; in particular, they cannot be preceded by the article *ne*.

The only way for these verbs to fill the functions of a noun is by derivation, of which there are three kinds. One is a suffix *-ove* which derives nouns from (some) verbs:

Table 7. Suffixal derivation of some verbs into nouns

	verb	derived N ('V-ing')
'walk'	*tō*	*tō-ove*
'go'	*vën*	*vën-ove*
'sit'	*sag*	*sag-ove*
'sleep'	*mitir̄*	*mitir̄-ove*
'stay'	*toge*	*tog-ove*

(82) Yöywye ti-ke   ti   ne **sag-ove**   nome   mi kema.
      thanks DAT-2SG DAT ART sit-NMLZR POSS:2SG with 1ex:pl
      'Thanks for *having sat* with us.'
      (lit. 'Thank you for *YOUR SITTING* with us')   [EP3-04a]

The roots of Table 7 are rigidly verbs; and the derived forms ending in *-ove* are rigidly nouns.[32]

The second type of derivation applies to transitive verbs. When they are nominalised together with their (nominal) object, the result is a compound noun ⟨V *vë* N⟩ with an incorporated object. The linker *vë* is only found in this nominalising structure. Such a word was found in (81) above, analysed here under (83):

(83) r̃ak   ne suqe   → ne   r̃ak-**vë**-suqe
      make ART initiation   ART make-NMLZR-initiations
      'perform initiations' → 'the performing of initiations'   [Stories.048]

(84) oye   ne gengon toq   → ne   oye-**vë**-gengon-toq
      take ART food   holy   ART take-NMLZR-food-holy
      'receive holy food'   → 'the Eucharist'   [Stories.033]

The contrast is rather clear between the verbal lexeme on the one hand, and the nominalised compound on the other hand, which behaves like a noun. The complex morphology of said compound shows that we are not dealing with simple conversion.

Finally, the third type of nominalisation in Hiw involves reduplication – a morphological device particularly common among Austronesian languages. Many verbs form their *nomen actionis* by reduplication (Table 8). While the simplex form of the word is a pure verb, the reduplicated form is a Noun–Verb hybrid: it can be a verbal form contextually endowed with reduplication, as is common for verbs; or it can be a nominalisation, in which case it behaves like any (Weak) noun in the language.

This process is productive, but not entirely predictable. For example, *tenteno* can mean both 'learning' (due to productive derivation from *teno* 'learn') and 'song'; *teptëp* means both 'loving' and 'a present'. All this information is stored in the lexicon.

---

32. Interestingly, the word *r̃ekove* 'work' probably results historically from an ancient nominalisation of the verb *r̃ak* 'do, make' (*r̃ak-ove* 'making' > *r̃ekove* 'work'). If this hypothesis is correct, this would be a rare case of a deverbal noun which has later been itself reanalysed as a Verb – Noun hybrid.

Table 8. Some reduplicated forms are both verbs and nouns

Simple	Verb	Reduplicated	Verb	Noun
gon	'eat:TR'	*gengon*	'DUP~eat'	'food, meal'
ni	'drink'	*nini*	'DUP~drink'	'drink'
sawe	'dance'	*sewesawe*	'DUP~dance'	'dance'
vë	'weave'	*vevë*	'DUP~weave'	'weaving' (Example 60)
qōr̃	'dream'	*qōr̃qōr̃*	'DUP~dream'	'dream'
tëp	'love, give'	*teptëp*	'DUP~love'	'love; present'
vatego	'teach'	*vetvatego*	'DUP~teach'	'teacher'
yog	'marry'	*yegyog*	'DUP~marry'	'wedding'
teno	'learn'	*tenteno*	'DUP~learn'	'learning; song'

5.1.3 *The limits of lexical flexibility*

In sum, it would be quite incorrect to infer, from the handful of examples given in Table 6 above, that Hiw freely allows any verb to take up the syntactic slots of nouns in the sentence, or vice versa. As we saw, only a handful of words show this dual syntactic distribution as verbs and as nouns.

For the major part of the lexicon, a verb can be made compatible with nominal functions only if it undergoes a morphological process of nominalisation – using either *-ove*, *-vë-*, or reduplication. Nouns and Verbs form two quite separate word classes in Hiw. The few lexemes illustrated in *Table 6*, which can function as Nouns as much as Verbs, are the exception rather than the norm. Compared to a language like English (where lexical conversion {NOUN ↔ VERB} is much more frequent, albeit not systematic), Hiw can be described as *lexically rigid*: a language where lexical items are typically assigned one word class, and hardly more.

5.2 Noun–locative hybrids

Another sign that lexical flexibility is limited in Hiw is that only a few combinations are actually found between word classes. Besides the Nouns – Verbs we just saw, Hiw has one Noun–Adverb hybrid:

– *vegyaye* '(N.) sign, mark – (ADV.) distinctively'

I have found two words which qualify both as an Adjective and as a Noun:

– *vr̃iwane* '(ADJ.) funny – (N.) fun'
– *meyigeyige* '(ADJ.) black, dark – (N.) darkness, Heathen times'

and one word which is a Verb, an Adjective and a Noun:

– *yöymer̃ën* '(N.) knowledge – (ADJ.) knowledgeable – (v.) know'

The only remaining combination that shows any form of frequency is a set of words that are both locative adjuncts [§4.6] and (Weak) nouns. These *noun–locative hybrids* include *tētēnēn* '(on) the beach'; *yo* '(on) the shore'; *metëkñaye, wōnaye* '(on) the road'; *yöte* '(in) the garden'; *yöye* '(in) the cave'. All place names, like *Hiw* 'Hiw island' or *Ostrëlye* 'Australia', also behave like *noun–locative hybrids*.

When these hybrid lexemes are used in a referential function, they behave like ordinary nouns in a noun phrase. They combine with the nominal article *ne* in order to form an argument phrase; they can take modifiers, etc.

(85) *Ne yöye in r̄ẽptog p' ëne.*
    ART cave ANAPH close FOC DEM
    [NOUN] 'That cave is just over there.'   [d01.15]

(86) *Ne yöte =nome në pusune.*
    ART garden =your STAT numerous
    [NOUN] 'You have many gardens.'   [Stories.040]

(87) Ike vati kema i *ne wōnaye*.
    2SG show 1EX:PL OBL ART road
    [NOUN] 'Show us the road.'   [Meravtit.194]

By definition, Noun–Locative hybrids can fill all the functions devoted to nouns, just like any other (Weak) noun. This comes in contrast with pure locatives (e.g. *yöñwe* 'at home'), which are not compatible with these nominal functions.

These same hybrid lexemes can also fill all the functions reserved to locatives [§4.6]: in particular, they can form syntactic adjuncts directly, with no need of a preposition. Thus, whereas a pure noun can only form a locative phrase by means of a preposition *yö* – as in (28) or (79) above – Noun-Locative hybrids do so directly:

(88) Ike yō ne yiwe ⟨mone r̄ög⟩ *yöye*.
    2SG see ART (arrow) lie:PL cave
    [LOCATIVE] 'You can see the arrows lying down *in the cave*.'   [Stories.008]

(89) Ike ⟨va ti⟩ *yöte* =nome?
    2SG plant:PL PAST garden your
    [LOCATIVE] 'Did you plant them *in your garden*?'   [d09.Karen:07]

(90) Kema ⟨teu r̄i-se⟩ *wōnaye*.
    1EX:PL meet-3PL road
    [LOCATIVE] 'We met them *on the road*.'   [d01.01]

Their status as Locative also explains why these words can combine with the 'originative' particle *te* [cf. (52) in §4.1], which is only compatible with locatives and not nouns:

(91) ne gengon te *yöte*
    ART food ORIG garden
    [LOCATIVE] 'food from the garden'

Whether a word will be a pure noun, a pure locative, or a hybrid, cannot be predicted from its semantics. For example, the lexeme *vönyö* 'island, country', despite referring to a location in space, is a pure noun, and not a locative. It needs a preposition to be transformed ("transferred", cf. 4.7.8) into a locative phrase:

(92) Sör̄ö ⟨tō ñwuye⟩ *yö vönyö* =sa.   (*tō ñwuye vönyö =sa)
    3DU go:NPL return LOC island their
    'They returned *to their island*.'   [Grouper.22]

(93) ne gengon te *yö vönyö* =ma
    ART food ORIG LOC island our
    'food from our island'   [d07.Kenu:01]

Table 9 compares the syntactic distribution of Noun–Locative hybrids with that of pure locatives on the one hand, and pure nouns on the other hand. (The table includes minor functions which have not been discussed in this study.) The list of functions open to hybrid lexemes is simply the union (in the mathematical sense, 'A∪B') of the functions associated with each of the individual word classes involved.

Table 9. The syntactic functions of NOUNS, LOCATIVES, and Noun–locative hybrids

Function	(Weak) nouns	Noun–locative hybrids	Lexical locatives	Locative preposition + weak noun
modifier of head in ArgP	+	+	–	–
direct possessor	+	+	–	–
associative modifier (with *i*)	+	+	–	–
[+ART] form an argument phrase	+	+	–	–
locative adjunct	–	+	+	+
locative topic	–	+	+	+
originative modifier (following *te*)	–	+	+	+
Examples	*vönyö* 'island'	*yöte* '(in) garden'	*yöñwe* 'at home'	*yö vönyö* 'in the island'

Such cases of Noun–Locative hybrids may be compared with some lexemes in English which also share syntactic properties both with nouns and locatives: see the example of *home* in (4) [§2.1]. Again, nothing makes it possible to predict that *home* can behave like an adjunct (e.g. *There's nobody home.*) whereas *house* cannot (**There's nobody house.*). Whether for English or for Hiw, the assignment of lexemes to word classes, and their status as hybrids, must be learned on a case-by-case basis, within the lexicon.

In sum, the system of Hiw is one in which a given lexeme is typically assigned one word class, and only one. Cases of multicategoriality are attested for sure: we just saw that some lexemes can function both as a verb and a noun; and others, both as a noun and a locative adjunct. However, these hybrid words are rather exceptional, and concern not more than a couple dozen lexical entries. The flexibility that characterises Hiw belongs not in the lexicon, but in the grammar, and in the multifunctionality assigned to its word classes. This is how Hiw can be described as grammatically flexible, but lexically rigid.

6. Cross-linguistic prospects

Following this detailed study of a single language, I propose to end the present study with a theoretical and methodological discussion, adopting this time a cross-linguistic perspective.

I will be addressing two questions. First, how can we apply the reasoning described here for Hiw to other languages with a different lexical and grammatical system? Second, assuming we can carry out distinct assessments of lexical flexibility for each language separately, is it possible to bring results together so as to compare flexibility across languages? What tools could help us build a cross-linguistic typology of lexical flexibility?

6.1   Analysing each language on its own grounds

One of the important principles of modern linguistics, as defined by Boas (1911) and Saussure (1916), is the realisation that every language ought to be understood based on a language-internal analysis. As the saying goes, each language must be "described in its own terms" (cf. Dryer 2006: 211, Haspelmath 2015).

The necessity of such a system-internal approach has been central to our own reasoning here. For example, we might observe superficially that two languages allow some of their nouns to head a predicate phrase; yet this information alone is not sufficient to start a meaningful cross-linguistic comparison. In some languages, the possibility to have nouns in predicative position is specific of just a few lexemes, for which the best analysis will be in terms of multicategoriality, i.e. a conversion from a noun to a verb [cf. English *ship* in (3a–c)]. In other languages, by contrast, the predicative function will be freely open to all nouns, to such an extent that it should even be included among the syntactic functions assigned to nouns in that system [cf. Hiw *maṝenage* in (7b)]. The former case takes place in the lexicon, the latter in the grammar: these are two distinct types of flexibility, which cannot be simply equated across languages based on similar-looking surface structures.

It is well known that every language cuts up the grammatical space, as it were, into a different set of categories and word classes (cf. Lemaréchal 1989; Hengeveld 1992; Haspelmath 2007, 2012, 2015); and the criteria that define, for example, *nouns* or *adjectives* in one language will be quite distinct in another language. And because languages can differ virtually on any aspect of these categories – on what lexemes they include, what syntactic functions they can fill… – one cannot just compare surface patterns across languages: a given construction must be situated system-internally, within the economy of its word classes.

Such a position provides strength to a language-specific analysis; but it becomes problematic as soon as we wish to apply the findings of one system to another one, or to start making comparisons. My view is that we should embrace this structuralist view – whereby each system needs to be described in its own terms – and analyse the economy of each system separately; yet this should not prevent us from defining a unified approach for that endeavour. The following pages propose to outline such an approach.

6.2   A unified method for assessing lexical flexibility

6.2.1   *Selecting relevant syntactic functions*
For each language under consideration, the first step is to identify a list of syntactic functions relevant to the language's grammar.

Some of these functions may be so common worldwide as to be potentially universal: {*head of* TAM *predicate*}, {*head of argument phrase*}, {*modifier of head in argument phrase*}… However, universal relevance need not be a criterion: in principle, the list of functions to be considered can also include language-specific structures, such as {*second element in a serial construction*}, {*incorporated object*}, {*host of suffix X*}… Any grammatical behaviour, no matter how idiosyncratic, is potentially useful to the language-internal assessment of grammatical functions, and of the word classes that map onto them.

The list of syntactic functions relevant to the study of lexical flexibility does not need to be comprehensive, or even long. Thus, in the context of this study of Hiw, I chose to consider six main functions.[33] A more fine-grained analysis could have included more of them (as suggested by Table 5 in §4.7.8, or Table 9 in §5.2), which would have refined the results. The only requirement is that the list of functions, regardless of its length, should be later used as a grid to assess all word classes.

Each function identified in the list should be defined in syntactic and functional terms, in such a way to avoid resorting to word classes. Thus, ethnocentric labels

---

33. See the left column of Table 1, §4.1, based on the functions examined in Section 3.

such as "verb phrases" and "noun phrases" ought to be replaced by neutral terminology, for instance "predicate phrase" or "argument phrase".[34] This is a cautionary step against making unnecessary assumptions about which functions can be filled by which word classes.

6.2.2 *Identifying word classes*

The second step is to identify a set of word classes. Languages tend to have major word classes that encompass most of the lexicon, as well as minor classes for particular paradigms (directionals, prepositions, conjunctions…). The inventory of parts of speech to be considered should at least include the major ones.

Again, there should be no attempt at identifying so-called universal parts of speech, such as "nouns" and "verbs". Because "pre-established categories don't exist" (Haspelmath 2007), the word classes to be defined will be different for each language anyway. The Adjectives of Hiw are a distinct kind of entity from the Adjectives of English, and we should not tie our hands in assuming that they should have any property in common; in this sense, the names given to word classes are but convenient labels (Haspelmath 2007, 2012). Some languages may lack a separate category of adjectives altogether.

As a corollary, it is perfectly acceptable to list parts of speech that appear to be unique to that language. In the case of Hiw, we saw that there was no such category as "noun" with a single grammatical behaviour, but two watertight classes, dubbed Strong Nouns vs. Weak Nouns. Similarly, distributional evidence showed that Hiw distinguishes neatly two sets of words, labelled Adverbs and Adjuncts [§3.3.2], even though other languages might have treated them as a single class of adverbs.

Each word class is defined as a set of lexical items which are empirically found to behave in an identical way in the language. The assignment of a given lexeme to a certain word class must be based strictly on its distribution in the corpus, rather than based on its meaning or its translation. For example, because *yeqën* means 'woman' and *yeqën tamesō* means 'old woman', it would be tempting to infer that *tamesō* must simply be an adjective meaning 'old', just like its English counterpart. However, a syntactic test shows that *tamesō* is compatible with the noun article [(63) *ne tamesō*], unlike Adjectives in Hiw: hence, this lexeme is not an adjective meaning 'old', but a (Weak) Noun better glossed 'old person', and (65) *yeqën tamesō* is a sequence of two nouns.

Likewise, only syntactic sets can tell us that *keko* 'child' is a Strong Noun, while *megoye* 'child' is a Weak Noun. In sum, just like the list of relevant word classes has to be language-specific, similarly the assignment of each lexeme to its own part of

---

34. This methodological point was made in §3.1 for Hiw.

speech should not be inferred based on its translation or on aprioristic assumptions, but should be assessed empirically.

6.2.3 *Multifunctionality: From word classes to functions*

Once a list of word classes has been identified, it is time to start mapping them onto the array of syntactic functions identified earlier. This *grammatical mapping* (word class → functions) constitutes the first step when assessing the economy of parts of speech in a language. As for the *lexical mapping* (lexeme → word classes), it depends on the results of that first step, and therefore can only be done later [§2.4, 6.2.4].

The question to be asked is the following:

(94) grammatical mapping
Given a word class C, let lexical item *x* be any random member of C.
Among the grammatical functions present on the reference list, which ones can be filled by *x* directly, without extra morphological material?

For instance, looking at the word class Noun in English, we could observe the behaviour of the randomly picked example *paper*. As showed in §2.1 above, we would see that it can be {head of an argument phrase} (Example 2a), and {modifier of an argument phrase} (Example 2b). This suggests that these two functions can be accessed by all Nouns in this language.

Crucially, (94) specifies that *x* should be "any random member" of C. In terms of method, this means that one lexical example is not sufficient to assess the behaviour of C as a whole. If we had randomly picked the noun *home*, we would have had the impression that nouns can directly access the function of clause adjunct, as in Example (4c) "There's nobody *home*." This, however, would have been a false positive: another member of the class Noun, e.g. *ship*, would have been ungrammatical in that position (**There's nobody ship*). The grammaticality of (4c) is not due to the status of *home* as a Noun, but to its dual lexical mapping (*home* → {noun; adverb}); this is a peculiarity of this specific lexeme, which tells us nothing about the behaviour of Nouns in general, and therefore must be set aside for the moment.

The same precaution should guide us when dealing with unfamiliar languages: the test (94) should be carried out recursively with several members of word class C, until a common denominator is identified for all of them. Every time a function F is found to be compatible with some members of C but incompatible with others (like that of adjunct for *home* but not for *ship*), then it should not be included in the list of functions regularly assigned to C.

Of course, not all combinations will be found in a corpus. For example, not all nouns are attested in all positions. This is when elicitation becomes necessary. Also important is the contrast between a combination that is unattested yet grammatical,

and one which is rejected by speakers. Thus, even if a large corpus of English will not include all nouns in a position to modify another noun, a phrase $\langle N_{modifier}\ N_{head}\rangle$ will always be grammatically acceptable, albeit pragmatically strange (ᵈ*a shoe hat*, ᵈ*a wheel cake*…). The grammaticality of these phrases, and the fact that native speakers can at least try to interpret them, makes it legitimate to assign that particular function (N_{modif}) to the entire class of Nouns: in English, it is a grammatical property of any noun that they can modify another noun. This same reasoning guided me when assigning to the nouns of Hiw the function {*head of* TAM *predicate*} [§4.7.7]: even though my corpus has actually few nouns in that position, attempts to elicit random nouns always brought about clauses that were accepted as grammatical and interpretable, if only semantically contrived.

The latter cases contrast with ungrammaticality. A sentence like **There's nobody ship* is ungrammatical in English, and will be rejected by speakers. This is not a matter of interpretability: it is in fact not too difficult to guess it might intend to mean *There's nobody on board this ship*. The sentence is simply incorrect in standard varieties of English, and this observation constitutes evidence when assessing grammatical mapping in a system. This allows us to rule out "adjunct" from the functions accessible to English Nouns.

Once all tests are carried out, the grammatical mapping of a given language can be presented in the form of a chart (see Table 1, §4.1 for Hiw). Minimally, each word class should be assigned at least one syntactic function. If a class C can access more than one function, then we have a case of grammatical flexibility, or *multifunctionality*. This paper claimed that multifunctionality is high in Hiw. Section §6.3 will propose a method for quantifying multifunctionality and compare it across languages.

6.2.4   *Multicategoriality: from lexemes to word classes*

We reach now the final step in analysing flexibility in a language: the observation of the *lexical mapping* from lexemes to word classes.

If a given lexeme is a well-behaved (non-hybrid) member of word class C, it should be automatically compatible with all the functions compatible with C, and exactly those: neither more nor less.

(95)   LEXICAL MAPPING
       A lexical item *x* is a member of a word class C *iff* it can occupy all the syntactic functions assigned to C

If we were to test the Hiw word *mesō* 'large', we'd see that it can fill exactly the four functions assigned to adjectives in this language (§4.3), and none other: it thus qualifies as a regular member of the Adjective class.

If a lexeme is attested in more functions than those normally assigned to its putative class C_1, then two cases are possible:

–   the item has been wrongly assigned to C_1, and should be reassigned to a different word class C_x;
–   or the item has been correctly assigned to C_1, but it is a hybrid (multicategorial) lexeme that combines two classes C_1 and C_2.

To illustrate the first scenario, let us take the example of Hiw *mët*, with a preliminary gloss 'die'. This word is compatible with all the functions normally associated with verbs, including {head of TAM predicate}. Examples like (96–96′) suggest that *mët* could probably be assigned the word class Verb with a meaning 'die', in a way parallel to *sō* 'fall':

(96)   Nine   ***mët***   ti.
       3SG   die   PAST
       'He died.'

(96′)   Nine   ***sō***   ti.
        3SG   fall   PAST
        'He fell.'

However, that same form *mët* is also commonly found as {*modifier in an argument phrase*}, a function which in Hiw is open to adjectives, but never to verbs:

(97)   ne   mon   ***mët***
       ART   bird   die/dead
       'a dead bird'

(97′)   *ne   mon   *sō*
        ART   bird   fall
        *a fallen bird

One option could be to say that *mët* is a multicategorial, "hybrid" lexeme: it would be a Verb 'die' in (96), but an Adjective 'dead' in (97). However, this analysis would be unnecessarily costly. Indeed, all the functions available to Verbs in Hiw are also available to Adjectives, even though the reverse is not true (Table 1, §4.1). The most economical analysis is to conclude that *mët* is always an Adjective (better glossed 'dead' in English), and never a Verb. Given the behaviour of Adjectives in Hiw [§4.3], this status is sufficient to account for all its occurrences in discourse, including in a sentence like (96).[35] There is no need to posit any conversion or

---

35. François (2003: 48–53) discusses this point in detail for the neighbouring language Mwotlap.

multicategoriality here: we are simply dealing with the multifunctionality inherent to the word class Adjectives in this language. The linguist should always aim for the most economical analysis, and posit conversion only if necessary.

The second scenario involves hybrid (i.e. multicategorial) lexemes. I have proposed that this is essentially the same as the relationship of conversion or heterosemy. Multicategoriality takes place when a given word combines all the properties of one word class C_1 with those of another word class C_2:

(98) MULTICATEGORIALITY
A lexical item belongs to two word classes C_1 and C_2 *iff* it can occupy, without any derivation, all the syntactic functions assigned to C_1 as well as those assigned to C_2.

Examples from English included *ship* and *home*, each with a dual affiliation, respectively as {Noun–Verb} and as {Noun–Adverb} [§2.1]. As for Hiw, we examined the case of {Noun–Verb} [§5.1] and {Noun–Locative} hybrids [§5.2]. A lexeme can belong to more than two word classes.

The fact that a particular word is multicategorial, is generally arbitrary and non-predictable: why is *home* a {Noun–Adverb} hybrid, but not *house*? Why is *ship* a {Noun–Verb} and not *car*? Likewise for Hiw, why is *veröye* 'fight' a {Noun–Verb}, but not *mañe* 'talk'? Equally non-predictable is the precise semantic shift between the two (sub)lexemes involved in a relation of conversion. The exact meaning of *ship* [v.] could not be inferred from its nominal meaning, and must be learnt separately. Likewise for the pair *pyë* [v.] 'fix bait' vs. *pyë* [n.] 'bait'. This is, again, an argument for considering that conversion, or multicategoriality, takes place in the lexicon rather than the grammar [§2.2].

While multicategoriality is high in English, I have argued that it is low in Hiw: in this language, most lexemes are assigned one word class, and only one. Section §6.3.4 will suggest a possible metric to quantify multicategoriality in a language.

6.3   Language comparison and quantification

Let me now come to the final question of cross-linguistic comparison.

One first approach could venture comparisons of substance between languages. For example, we could search for which languages allow nouns to head a TAM predicate (cf. Nordlinger & Sadler 2004), or verbs to head an argument. In some languages, such patterns may be better analysed as conversion or multicategoriality (NOUN → VERB, or VERB → NOUN). In other languages, they may be better described as multifunctionality, as I proposed for Hiw: nouns remaining nouns even when predicative; verbs remaining verbs even when used as arguments. This is a delicate

matter which ought to be undertaken carefully for each language, looking at the key criteria of universality and semantic compositionality (cf. Evans & Osada 2005).[36]

Among the various approaches possible, I will briefly outline a method for quantifying word-class flexibility in each language, by measuring (multi)functionality on the one hand – via two different metrics – and (multi)categoriality on the other. These metrics could be calculated language-internally, and then be compared across languages.

6.3.1   *A metric for multifunctionality*

The degree of multifunctionality in a grammatical system can be visualised in the form of a chart, such as Table 1, §4.1. More than just a visualising tool, such a table can be used as a base for assessing multifunctionality as a metric. Indeed, it allows us to count the exact number of syntactic functions that the language licenses for each word class.

For the reader's comfort, Table 1 is here reproduced as Table 10.

Table 10. Major word classes in Hiw and their syntactic functions: a matrix for a metric

Syntactic function	Verb	Adjective	Numeral	Strong noun	Weak noun	Adverb	Adjunct
head of argument phrase	–	–	+	+	–	–	–
modifier in argument phrase	–	+	+	–	+	–	–
head of TAM-inflected predicate	+	+	+	+	+	–	–
head of direct predicate	+	+	+	+	–	–	–
modifier in predicate phrase	+	+	–	–	–	+	–
adjunct	–	–	–	–	–	–	+

Given seven major word classes considered there, a minimal mapping – the one found in a hypothetical language showing extreme rigidity – would have been seven, each word class being assigned a single function. The maximum possible mapping would be 7 (word classes) times 6 (main functions), namely 42; this number would correspond to a hypothetical language that would be maximally multifunctional. Between these two canonical extremes, Hiw shows a rate of 18 matches (the number of '+' signs in *Table 10*), i.e. an average of $f = 18/7 = 2.57$ syntactic functions for each word class. This could be labelled a rate of MULTIFUNCTIONALITY f. Compared to a maximally rigid language, for which f would equal 1 (one function per word class), the multifunctionality score of Hiw ($f = 2.57$) is relatively high.

The general formula for calculating a language's multifunctionality score *f* is given in (99):

(99) MEASURING MULTIFUNCTIONALITY

For a given language, let κ (kappa) be the number of major word classes, and σ (sigma) the number of main syntactic functions examined. The grammatical mapping of word classes onto functions yields a number M of positive matches (i.e. sum of all the functions regularly activated by the various word classes). That number M is such that $\kappa < M < (\kappa \times \sigma)$. The multifunctionality score *f* indicates how many functions, on average, are tied to each word class, and is defined as:

$$f = \frac{M}{K}$$

6.3.2 A metric for grammatical flexibility

Remaining on the grammar side, we might want to describe grammatical flexibility as the manner in which a given system has taken advantage of the available "grammatical space". Knowing that $\{ \kappa < M < (\kappa \times \sigma) \}$, we can consider that the available space, what we could call the "maximum leeway" for flexibility, corresponds to $[(\kappa \times \sigma) - \kappa]$.

In the case of Hiw, $[(\kappa \times \sigma) - \kappa] = 42 - 7 = 35$. That is, once each word class has been assigned one function (one '+' sign per column in the table), there are 35 boxes left in the table. The more '+' signs are found in these 35 boxes, the more flexible the language. Now, out of these 35 available class/function pairings, the grammar of Hiw instantiates $(M - \kappa) = (18 - 7) = 11$. This yields a rate of 11/35 = 31.4%: when assigning functions to its word classes, Hiw takes advantage of 31.4 percent of the available grammatical space.

Let us call this second metric a score of GRAMMATICAL FLEXIBILITY, abbreviated φ (phi):

(100) MEASURING GRAMMATICAL FLEXIBILITY

For a given language, let κ (kappa) be the number of major word classes, and σ (sigma) the number of main syntactic functions examined. The grammatical mapping of word classes onto functions yields a number M of positive matches. That number M is such that $\kappa < M < (\kappa \times \sigma)$. The grammatical space available for flexibility is measured as $[(\kappa \times \sigma) - \kappa]$. The score φ of grammatical flexibility is defined as:

$$\varphi = \frac{(M - \kappa)}{[(\kappa \times \sigma) - \kappa]}$$

Should similar measurements be carried out for other languages, Hiw would probably rank comparatively high. Obviously, cross-linguistic comparison would take us beyond the scope of this study; yet for a quick comparison, we might look at the very rough, admittedly simplified representation of the grammatical mapping of English proposed in §2. Figure 1, crossing $\kappa = 4$ major parts of speech and $\sigma = 4$ main functions, showed only $M = 5$ matches: this yields a multifunctionality rate of $f_{\text{ENG}} = (M/\kappa) = 5/4 = 1.25$ syntactic function per word class, significantly lower than the 2.57 rate for Hiw.

As for the potential for grammatical flexibility there, it equalled $[(\kappa \times \sigma) - \kappa] = (16 - 4) = 12$ (that is, twelve class/function pairings available for flexibility besides the four minimal ones). Out of these, English apparently has only taken advantage of $(M - \kappa) = 5 - 4 = 1$ such extra pairing. This yields a grammatical-flexibility rating of $\varphi_{\text{ENG}} = 1/12 = 8.3\%$: again, way lower than the 31.4% of Hiw.

6.3.3 A note on Tahitian

Jacques Vernaudon (n.d.) is a report on the word classes of Tahitian, the language of his expertise.[37] Following the method exposed in §6.2 above, he built *Table 11*.

Table 11. Major word classes in Tahitian and their syntactic functions

Syntactic function	Verb	Adjective	Numeral	Strong noun	Weak noun	Adverb	Locative
Example	*'amu* 'eat'	*rahi* 'large'	*piti* 'five'	*'aiū* 'baby'	*vahine* 'woman'	*ato'a* 'all'	*ananahi* 'yesterday'
argument phrase	–	–	–	+	–	–	+
modifier in argument phrase	+	+	+	+	+	–	–
TAM-inflected predicate	+	+	+	+	+	–	–
equational predicate	+	–	–	+	+	–	+
attributive predicate	+	+	–	+	+	–	–
locative predicate	–	–	–	–	–	–	+
modifier in predicate phrase	+	+	–	+	+	+	–
adjunct	–	–	–	–	–	–	+

For reasons of space, I won't be able to go into the detail of Vernaudon's demonstration, and will only display his results. This valuable data will enable us to calculate the scores f_{TAH} and φ_{TAH} for Tahitian. Crossing $\kappa = 7$ major parts of speech and $\sigma = 8$ main functions, the number of positive matches found is $M = 27$. This

---

37. Vernaudon (n.d.). On related questions, see Vernaudon & Rigo (2004); Paia & Vernaudon (2004); Vernaudon (2011).

yields a score of (multi)functionality $f_{\text{TAH}} = M/\kappa = 27/7 = 3.86$. As for grammatical flexibility, we get:

$$\varphi_{\text{TAH}} = \frac{(M-\kappa)}{[(\kappa \times \sigma) - \kappa]} = \frac{(27-7)}{[(7 \times 8)-7]} = \frac{20}{49} = 40.8\%$$

Of course, it would be desirable to delve into the detail of each language cited, in the same fashion as we did for Hiw in the present study. But let us assume, for the sake of discussion, that the data and the reasonings summarised here are correct, for the three languages mentioned. The results shown in *Table 12* suggest that languages can be compared with respect to their degree of multifunctionality (*f*) and grammatical flexibility (*φ*). Interestingly, Hiw appears to be much more grammatically flexible than English, but is still outranked by Tahitian.

Table 12. Two metrics for measuring multifunctionality and grammatical flexibility: a cross-linguistic (preliminary) comparison

	Multifunctionality (*f*)	Grammatical flexibility (*φ*)
English	1.25	8.3 %
Hiw	2.57	31.4 %
Tahitian	3.86	40.8 %

6.3.4   *A metric for multicategoriality*

We just saw how flexibility could be measured in the grammar component of a language. But what about its lexical component? Can we also measure multicategoriality?

The study of grammar makes it possible to bring together data in the form of a synthetic chart such as *Table 10* above, and to propose calculations on this basis. The lexicon seems more difficult to explore, due to the sheer number of specific lexemes to be considered. However, the task is not impossible, assuming we can count on lexicographic resources. Multicategoriality could then be assessed, quite simply, by counting the number of lexemes associated with more than one part of speech, as a proportion of the whole lexicon. More fine-grained measurements could count the proportion, say, of *noun–verbs* or *noun–adverbs* within the Noun lexicon.

The dictionary of Hiw is still in progress. Out of a larger corpus of handwritten texts, I have transcribed 14 texts in electronic form, totalling 17,600 words. That corpus contains 1345 lemmas, yielding a preliminary lexicon of as many entries. These include 1188 proper lexemes (verbs, adjectives, nouns, adverbs, adjuncts, numerals), and 157 other entries (pronouns, affixes, TAM markers, interjections, etc.). This resource does not include all the lexical data I have collected (for example, it lacks most fauna and flora terms, elicited separately), but it constitutes a representative sample of the language. Combing this lexicon revealed that, among the 1188

lexical entries, only 28 were multicategorial:[38] 18 noun–verbs, 6 noun–locatives, and 4 miscellaneous (see Section 5).

Using a simple calculation, we can conclude that the lexicon of Hiw shows a very low number of multicategorial lexemes: 28/1188 = 2.36 %. The remaining 1160 lexemes (97.64%) are "rigid" lexemes, i.e. lexical items assigned to only one word class. I do not have similar numbers for English; but given appropriate lexicographic resources (ideally digital) these should not be too difficult to obtain. It is a safe bet to say that English would rank much higher than Hiw in terms of multicategoriality.

Tahitian can provide a useful example of the sort of measurement that could be done, based on actual publications. This language has the advantage of being described in various lexicographic resources (cf. Charpentier & François 2015), and this includes the very thorough dictionary created by the Tahitian Academy (Académie Tahitienne 1999). The latter dictionary now exists in electronic format.[39] Gaining access to the source files, Jacques Vernaudon was able to conduct some observations (Vernaudon n.d.).

Out of a total 12,297 lexical entries, 546 (i.e. 4.4%) constitute grammatical morphemes; we want to concentrate on the 11,751 remaining entries, which are proper lexemes. The Académie's dictionary has assigned to each of these lexemes one or more among four parts of speech: there are 7481 entries identified as (common) nouns; 1881 as adjectives; 4909 as verbs; 211 as adverbs. If we add up all these lexical word-class assignments, we reach a total of 14,482. If we use the terms I proposed in §2.2, we could say that the dictionary has 11,751 lexemes, and 14,482 "sublexemes" (=pairs linking a form and a word class).

Thanks to this data, we can measure multicategoriality in the lexicon in the following way (after Vernaudon, n.d.):

(101)   MEASURING MULTICATEGORIALITY
   For a given language endowed with reliable lexicographic resources, let *E* be the total number of lexical entries (number of "\lx"), and *P* the number of sublexemes identified by each word class (number of "\ps"). Each \lx entry must have at least one \ps, but hybrid entries have more than one; hence *P*>*E*. We can calculate the rate μ ('mu') of multicategoriality in the lexicon, defined as:

$$\mu = \frac{P}{E}$$

---

38. I have not counted as multicategorial those verbs which showed lability (Letuchiy 2009) between an intransitive and a transitive use: these were counted simply as Verbs.

39. http://www.farevanaa.pf/dictionnaire.php

In a purely rigid language – one in which each lexeme would have one word class and only one – we would have P=E, hence μ=P/E = 1. In a dictionary with only four lexical parts of speech (noun, adjective, verb, adverb), a maximally multi-categorial lexicon would be one in which each entry is assigned all four classes, hence μ = (4n/n) = 4. The multicategoriality of Tahitian therefore necessary ranges between $\mu = 1$ and $\mu = 4$. The actual score we find is:

$$\mu_{\text{TAH}} = \frac{P}{E} = \frac{14482}{11751} = 1.232$$

This means that, on average, there are 123 word-class assignments (or 123 sublexemes) for every 100 lexical entries: this suggests that lexical conversion is rather limited in Tahitian.

Using the same calculation for Hiw (though with a less elaborate lexicon, and rougher figures) we find that $E = 1188$, and $P = 1217$.[40] So $\mu_{\text{HIW}} = 1217/1188 = 1.024$.

It seems fair to conclude that both Tahitian and Hiw have low scores of multi-categoriality. Whatever flexibility exists in these languages resides not so much in their lexicon, but owes much more to the grammatical adaptability of their word classes.

7.   Conclusion: Lexically rigid, grammatically flexible

This study has proposed to break the notion of lexical flexibility into two separate operations. One mapping, from lexemes onto word classes, takes place in the lexicon, and has its information stored under each lexical item. The second mapping, from word classes onto syntactic functions, has its locus in the grammar. These two operations differ in nature, and in their impact on the system. They are in principle independent from each other: a language may be characterised by a rigid, one-to-one mapping in the lexicon, but a more flexible – one-to-many – mapping in the grammar; or vice versa; or it can be rigid on both sides, or flexible on both sides.

Through a detailed examination of the Hiw grammar, I have proposed that its relatively high flexibility resides mostly in its grammar component. The economy of word classes in this language is one where each lexical item is typically "unicate-gorial" – being assigned just one part of speech; but then, each of these word classes has access to a relatively large array of syntactic functions. We saw, for example, that Hiw allows most of its word classes (verb, adjective, numeral, nouns…) to head a TAM-inflected predicate, a function which would be reserved to verbs in a language

like English. Likewise, Numerals are found in at least five syntactic functions (*head of argument phrase; modifier in argument phrase; head of TAM predicate; head of direct predicate; modifier in predicate phrase*). I described this form of flexibility in the grammar with the concept of multifunctionality. I have also proposed ways to quantify it, so as to compare this dimension across languages.

It would be instructive to see how far this analysis could be extended to other Oceanic languages, and assess whether Hiw is typical of its family, or not. And beyond the Pacific, future studies may want to examine how these two dimensions (multicategoriality in the lexicon, multifunctionality in the grammar) are treated in the language families of the world. Of course, every system has its own inventory of parts of speech, and should be described in its own terms; languages also differ as to which syntactic functions are most relevant in contrasting word classes. While the methodology proposed here does take the structural uniqueness of each system seriously, it also includes a number of options to compare degrees of flexibility across languages. Future cross-linguistic studies could help us understand how the world's languages handle the potential of their word classes, and see to what extent they take advantage of the grammatical space at their disposal.

Abbreviations

ANAPH	anaphoric	INTSF	intensifier
APPREH	apprehensional modality	IPFV	imperfective
ART	article	LOC	locative
BKPF	Background Perfect	MX	mixed gender
CAUS	causal preposition	NMLZR	nominaliser
COMP	complementiser	NPL	non-plural (verbal number)
CONT	continuative aspect	OBL	oblique
CPLT	complete aspect	ORIG	originative
DEM	demonstrative	POSS	possessive classifier or linker
DISTR	distributive	POT	potential
DOM	differential object marking	REL	relativiser
FOC	focus	SBJV	subjunctive
FUT	future	SEQ	sequential aspect
HUM	gender classifier for humans	STAT	stative aspect
INDF	indefinite	TAM	Tense-Aspect-Mood

Hiw orthography

orth.	a	e	ë	ē	g	i	k	m	n	ñ	ñw	o	ö	ō	p	q	r̄	s	t	u	v	w	y
IPA	a	ə	ɛ	ɪ	ɣ	i	k	m	n	ŋ	ŋʷ	ɔ	ɵ	o	p	kʷ	ᵍʟ	s	t	ʉ	β	w	j

---

40. This corresponds to the number of entries E, plus 29 extra \ps fields: 27 hybrid lexemes showing two \ps, plus one that has not two but three word classes (*yöymer̄ën*, see §5.2).

References

Abney, Steven Paul. 1987. *The English noun phrase in its sentential aspect*. Doctoral dissertation. Massachusetts Institute of Technology.

Académie Tahitienne. 1999. *Dictionnaire Tahitien-Français – Fa'atoro Parau Tahiti-Fārani*. Papeete: Fare Vāna'a. (online: http://www.farevanaa.pf/dictionnaire.php)

Aikhenvald, Alexandra. 1999. Serial constructions and verb compounding. Evidence from Tariana (North Arawak). *Studies in Language* 23 (3), 471–497. doi:10.1075/sl.23.3.02aik

Aikhenvald, Alexandra. 2006a. Serial verb constructions in a typological perspective. In Dixon & Aikhenvald (eds.), 1–87.

Aikhenvald, Alexandra. 2006b. Serial verb constructions in Tariana. In Dixon & Aikhenvald (eds.), 178–201.

Ameka, Felix K.; Alan Dench & Nicholas Evans (eds.). 2006. *Catching Language: The Standing Challenge of Grammar Writing*. (Trends in linguistics studies and monographs, 167.) Berlin: Walter de Gruyter. doi:10.1515/9783110197693

Baker, Mark. 1989. Object sharing and projection in serial verb constructions. *Linguistic inquiry* 20.4: 513–553.

Bauer, Laurie. 1998. When is a sequence of two nouns a compound in English? *English Language and Linguistics* 2.01 (1998): 65–86.

Benveniste, Émile. 1974 [1967]. Fondements syntaxiques de la composition nominale. *Problèmes de Linguistique Générale* 145–162. Paris: Gallimard.

Boas, Franz. 1911. Introduction. In Franz Boas (ed.), *Handbook of American Indian Languages*, 1–83. Washington, DC: Bureau of American Ethnology.

Bril, Isabelle & Françoise Ozanne-Rivierre (eds.), *Complex predicates in Oceanic languages: Studies in the dynamics of binding and boundness*. (Empirical approaches to language typology, 29.) Berlin: Mouton de Gruyter. doi:10.1515/9783110913286

Charpentier, Jean-Michel & Alexandre François. 2015. *Atlas Linguistique de la Polynésie Française – Linguistic Atlas of French Polynesia*. Berlin & Papeete: Mouton de Gruyter & Université de la Polynésie Française. (online: https://hal.archives-ouvertes.fr/hal-01120553v1) doi:10.1515/9783110260359

Crowley, Terry. 1982. *The Paamese language of Vanuatu*. Pacific Linguistics, B-87. Canberra: Australian National University.

Crowley, Terry. 2002. *Serial Verbs in Oceanic: A Descriptive Typology*. Oxford: OUP. doi:10.1093/acprof:oso/9780198241355.001.0001

Dixon, Robert M. W. & Alexandra Aikhenvald (eds.). 2006. *Serial Verb Constructions: A cross-linguistic typology*. Oxford: Oxford University Press.

Dryer, Matthew S. 2006. Descriptive theories, explanatory theories, and Basic Linguistic Theory. In Ameka, Dench & Evans (eds.), 207–234.

Durie, Mark. 1997. Grammatical structures in verb serialization. In Alex Alsina, Joan Bresnan & Peters Sells (eds.), *Complex predicates*, 289–354. Stanford: CSLI publications.

Enfield, Nicholas I. 2006. Heterosemy and the grammar-lexicon trade-off. In Ameka, Dench & Evans (eds.), 1–24.

Evans, Nicholas & Toshiki Osada. 2005. Mundari: The myth of a language without word classes. *Linguistic Typology* 9, 351–390.

Foley, William & Olson, Mike. 1985. Clausehood and verb serialization. In Johanna Nichols & Anthony Woodbury (eds.), *Grammar inside and outside the clause. Some approaches to theory from the field*, 17–60. Cambridge: Cambridge University Press.

François, Alexandre. 2000. Dérivation lexicale et variations d'actance: Petits arrangements avec la syntaxe. *Bulletin de la Société de Linguistique*, xcv:1, 15–42. doi:10.2143/BSL.95.1.503717

François, Alexandre. 2003. *La sémantique du prédicat en mwotlap (Vanuatu)*. Collection Linguistique de la Société de Linguistique de Paris, 84. Paris, Louvain: Peeters.

François, Alexandre. 2004. Chains of freedom: Constraints and creativity in the macro-verb strategies of Mwotlap. In Bril & Ozanne-Rivierre (eds.), 107–143.

François, Alexandre. 2005a. A typological overview of Mwotlap, an Oceanic language of Vanuatu. *Linguistic Typology* 9–1, 115–146.

François, Alexandre. 2005b. Diversité des prédicats non verbaux dans quelques langues océaniennes. In Jacques François & Irmtraud Behr (eds.), *Les constituants prédicatifs et la diversité des langues: Actes de la Journée de la Société de Linguistique de Paris*, 179–197. Louvain: Peeters.

François, Alexandre. 2006. Serial verb constructions in Mwotlap. In Dixon & Aikhenvald (eds.), 223–238.

François, Alexandre. 2009. Verbal number and Suppletion in Hiw. Paper presented at *11·ical*, Eleventh International Conference on Austronesian Linguistics, Paris, June 2009. (https://www.academia.edu/8403922)

François, Alexandre. 2010a. Phonotactics and the prestopped velar lateral of Hiw: Resolving the ambiguity of a complex segment. *Phonology* 27, 393–434. doi:10.1017/S0952675710000205

François, Alexandre. 2010b. Pragmatic demotion and clause dependency: On two atypical subordinating strategies in Lo Toga and Hiw (Torres, Vanuatu). In Isabelle Bril (ed.), *Clause hierarchy and Clause linking: the Syntax and pragmatics interface*, 499–548. Amsterdam, New York: Benjamins. doi:10.1075/slcs.121.16fra

François, Alexandre. 2011. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1, 175–246. doi:10.1075/jhl.1.2.03fra

François, Alexandre. 2012. The dynamics of linguistic diversity: Egalitarian multilingualism and power imbalance among northern Vanuatu languages. *International Journal of the Sociology of Language* 214, 85–110.

François, Alexandre. 2013. Shadows of bygone lives: The histories of spiritual words in northern Vanuatu. In Robert Mailhammer (ed.). *Lexical and structural etymology: Beyond word histories*. Studies in Language Change, 11. Berlin: DeGruyter Mouton, 185–244. doi:10.1515/9781614510581.185

François, Alexandre. 2014a. Trees, Waves and Linkages: Models of Language Diversification. In Claire Bowern & Bethwyn Evans (eds), *The Routledge Handbook of Historical Linguistics*. New York: Routledge, 161–189.

François, Alexandre. 2014b. The morphology of object marking in Hiw. Paper read at 7th *Austronesian and Papuan Languages and Linguistics* (APLL7), School of Oriental and African Studies (SOAS), London.

François, Alexandre. 2015. The ins and outs of *up* and *down*: Disentangling the nine geocentric space systems of Torres and Banks languages. In *The Languages of Vanuatu: Unity and Diversity*, ed. by A. François, S. Lacrampe, M. Franjieh & S. Schnell. (Studies in the Languages of Island Melanesia, 5.) Canberra: Asia Pacific Linguistics Open Access. 137–195.

François, Alexandre. in prep. *A grammar of Hiw, Vanuatu*. Unpublished ms.

Haspelmath, Martin. 2007. Pre-established categories don't exist: Cnsequences for language description and typology. *Linguistic Typology* 11, 119–132. doi:10.1515/LINGTY.2007.011

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86 (3): 663–687. doi: 10.1353/lan.2010.0021

Haspelmath, Martin. 2012. How to compare major word classes across the world's languages. In T. Graf, D. Paperno, A. Szabolcsi & J. Tellings (eds), *Theories of everything: In honor of Edward Keenan*. UCLA Working Papers in Linguistics, 17: 109–130.

Haspelmath, Martin. 2015. Framework-free grammatical theory. In Heine, Bernd & Heiko Narrog, eds. *The Oxford handbook of linguistic analysis*. Oxford: Oxford University Press, 287–310.

Hengeveld, Kees. 1992. Parts of speech. In Michael Fortescue, Peter Harder & Lars Kristoffersen (eds.), *Layered structure and reference in a functional perspective* (Pragmatics and Beyond New Series 23). Amsterdam: Benjamins. 29–55. doi: 10.1075/pbns.23.04hen

Hengeveld, Kees & Eva van Lier. 2010. An implicational map of parts of speech. *Linguistic Discovery*, 8 (1), 129–156. doi: 10.1349/PS1.1537-0852.A.348

Launey, Michel. 1994. *Une grammaire omniprédicative: Essai sur la morphosyntaxe du nahuatl classique*. Sciences du Langage, Paris: CNRS.

Lazard, Gilbert & Louise Peltzer. 2000. *Structure de la langue tahitienne*. (Langues et Cultures du Pacifique 15) Société d'études linguistiques et anthropologiques de France 391. Paris, Louvain: Peeters.

Lehmann, Christian. 2008. Roots, stems and word classes. *Studies in Language* 32.3: 546–567.

Lemaréchal, Alain. 1989. *Les parties du discours, Syntaxe et sémantique*. Linguistique Nouvelle. Paris: Presses Universitaires de France.

Lemaréchal, Alain. 1994. Désignation et dénomination : Superparties du discours et parties du discours. In Louis Basset & Marcel Pérennec (eds.), *Les classes de mots: traditions et perspectives*. Lyon: Presses Universitaires de Lyon. 149–168.

Letuchiy, Alexander. 2009. Towards a typology of labile verbs: Lability vs. derivation. In Patience Epps & Alexandre Arkhipov (eds.), *New challenges in typology: Transcending the borders and refining the distinctions*, 247–268. Berlin: Mouton deGruyter.

Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67 (3): 475–509. doi: 10.1353/lan.1991.0009

Nordlinger, Rachel & Louisa Sadler. 2004. Nominal tense in crosslinguistic perspective. *Language* 80: 776–806. doi: 10.1353/lan.2004.0219

Paia, Mirose & Jacques Vernaudon. 2004. Complex predicates in Tahitian: A particular case of qualitative modification. In Bril & Ozanne-Rivierre (eds.), 251–262.

Radford, Andrew. 1997. *Syntax: A minimalist introduction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139166898

Ryder, Mary Ellen. 1994. *Ordered chaos: The interpretation of English noun-noun compounds*. Berkeley: Univ. of California Press.

Saussure, Ferdinand de. 1972 [1916]. *Cours de linguistique générale*. Edited and presented by Tullio de Mauro, Paris: Payot.

Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.

Verhaar, John W. M. 1984. The Categorial System in Contemporary Indonesian: Verbs. In K. Purwo (ed.), *Towards a Description of Contemporary Indonesian: Part I*, 27–64. Jakarta: NUSA, Universitas Atma Jaya.

Vernaudon, Jacques. 2011. Grammaticalization of Tahitian *mea* 'thing, matter' into a stative aspect. In Claire Moyse-Faurie & Joachim Sabel (eds.), *Topics in Oceanic Morphosyntax*, 319–339. Berlin: Mouton de Gruyter. doi: 0.1515/9783110259919.319

Vernaudon, Jacques. n.d. Notes sur les parties du discours en tahitien. Unpublished notes (Jan 2017). Tahiti: University of French Polynesia.

Vernaudon, Jacques & Bernard Rigo. 2004. De la translation substantivante à la quantification: Vers une caractérisation sémantique de l'article *te* en tahitien. *Bulletin de la Société Linguistique de Paris* 99: 457–480. doi: 10.2143/BSL.99.1.541921

Veselinova, Ljuba. 2013. Verbal number and suppletion. In Matthew Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Chapter 80 (online: http://wals.info/chapter/80).

Author's address

Alexandre François
Centre National de la Recherche Scientifique, Langues et Civilisations à Tradition Orale
7 rue Guy Moquet
F-94800 Villejuif
France

alexandre.francois@cnrs.fr