

Vietnamese 'Morphology' and the Definition of Word

Rolf Noyer

1. Introduction

According to the Lexicalist Hypothesis, it is the responsibility of the Lexicon to generate the well-formed words of a language, where by 'word' what is meant is a structure which is 'opaque to all sentence-level operations and descriptions' (Di Sciullo & Williams 1987:52).¹ If this claim is to have any content however, it must be shown that a variety of independent criteria converge on the notion 'word' as distinct from any other syntactic structure. Recent work under the heading 'Distributed Morphology' (Noyer 1997, Halle & Marantz 1993) has questioned the existence of a clear-cut boundary between word-syntax and morpheme-syntax, returning to the assumptions of a pre-lexicalist generative syntax such as was found in *Syntactic Structures* (Chomsky 1957). The functions attributed to the Lexicon are in this theory distributed into various other modules of grammar, including a generalized (morpho)syntax, a component of autonomous Morphology largely concerned with readjustment rules and allomorphic choice, and an Encyclopedia associating idiomatic meanings with phonological forms (sometimes in specific environments).

In this paper I examine Vietnamese, a language normally thought devoid of morphology and for which the debate regarding the definition of 'word' was notoriously contentious in structuralist treatments (Thompson 1963). The criteria normally associated with wordhood are shown to apply to domains which are not syntactically opaque at all, considerably weakening the thesis of word atomicity.

Of the various criteria which normally identify 'words' as opposed to syntactic forms we can identify several of importance here. First, inasmuch as the lexicon produces 'words' and 'words' project syntax with a compositional semantics, we normally equate 'word' with sign, that is, with the domain of idiomaticity. Second, syntax is said to be 'productive' while morphology need not be. That is, the selectional restrictions holding among morphological constituents can be arbitrary (*London-er* vs. *Boston-ian*) while those holding among words are in some sense principled or systematic. Third, the word is typically the domain of morphophonological operations such as reduplication, hence *fuzzy-wuzzy* is one word, but *fuzzy animal* need

¹This paper was originally presented at the Conference on Lexical Structures, Wuppertal, August 28, 1995. I would like to thank Alec Marantz, Heidi Harley, and an anonymous reviewer for their comments. Thanks also to Sonny Vu for valuable discussion and for data judgments.

not be one word. Finally, any domain identified as a word from these criteria should be opaque to syntax.

Data from Vietnamese presented here show that none of the criteria mentioned above—idiomaticity, arbitrary selectional restrictions, morphophonology—correlates with ‘syntactic atom’ in any necessary way. Instead, the syntactic atom may be (1) meaningless in isolation, (2) unproductive in its composition and (3) morphophonologically related to other syntactic atoms via a morphophonological process, reduplication. There are no criteria converging on a unitary notion ‘word’ in Vietnamese. A consequence of this is that Vietnamese grammarians have disagreed strenuously about whether certain facts are to be labelled ‘morphology’ or ‘syntax.’ I propose here that the distinction, as a clear-cut dividing line, is vacuous.

The data in this paper come from published sources as well as native speakers. I have consulted two important studies of Vietnamese morphosyntax from a generative perspective, Phong (1976) and Nhân (1984). The latter is especially detailed in its classification and explication of the variety of reduplicative and compound structures. A native speaker consultant, Phúc Thị Ngọc Lê, provided patient and insightful assistance during the academic year 1994-1995. Finally, I have benefited from discussions and comments by Sonny Vu, whose recent work (1998a, 1998b) promises to be a significant expansion (and perhaps correction) of the ideas presented below.

2. Idiomaticity and Availability

It will be convenient to begin by defining two important notions: *idiomaticity* and *availability*. Consider the following pairs of expressions.

(1)	English	Vietnamese	
a.	<i>atlas</i> ‘book with maps’	<i>dèn</i> lamp	‘lamp’
b.	<i>per-cuss-ion</i> ‘drums, etc.’	<i>đế đô</i> emperor-metropolis	‘capital’
c.	<i>huckle-berry</i> ‘kind of berry’	<i>vườn tược</i> garden-?	‘gardens’
d.	<i>carpal tunnel syndrome</i> ‘pathology of the wrist’	<i>quốc phục</i> national costume	‘national costume’
e.	<i>hard-nose</i> ‘strict person’	<i>dèn sách</i> lamp book	‘to study’
f.	<i>take the veil</i> ‘become a nun’	<i>xoe tơ kết tóc</i> bind silk weave hair	‘get married’

Following the usage of Marantz (1995), I will use the term *idiom* to refer to any phonologically identified structure whose meaning is not predictable from the meaning of the subparts of the structure along with universal principles of interpretation of the structure. All the expressions in (1) are idioms in this sense, even *atlas* which is monomorphemic. (The meaning of *atlas* is not predictable from its subparts since *atlas* has no subparts.) Idiomaticity is strictly correlated neither with indivisible constituents ('morphemes') nor with words nor with phrases. Although, as Aronoff (1976) puts it, 'the word gravitates to the sign,' being a sign and being a word in fact have no necessary connection, and many signs are non-words in Vietnamese. For Vietnamese, and probably for all languages, it is impossible to dismiss idiomatic superword-sized syntagms as an exceptional excrescence on the lexicon.

I will use the term *available* (cf. Fr. *disponible* from Corbin 1987) to refer to an element whose presence in the string does not imply the presence of some member of an arbitrary list. English *per-*, *-cuss-*, *huckle-* and *carpal* are 'unavailable' in this sense (perhaps also *-ion*), since they combine with an arbitrarily limited set of forms; for example, *huckle-* combines only with *berry*, and for most speakers *carpal* is not an adjective referring to the wrist, but rather occurs only in the phrase *carpal tunnel syndrome*. The remaining terms are available, that is to say, for a given expression containing *veil* or *nose* one can make no inferences regarding a completely arbitrary set of other terms which must also occur in that expression.

Considering only idiomaticity and availability, there are no differences between the English expressions in (1) and the analogous Vietnamese ones. Contrary to common claims that Vietnamese is a canonically 'isolating' language in which the syllable and the morpheme or word are coextensive, in fact the syllable and the *idiom* as defined here are not normally coextensive in Vietnamese. Phong (1976), for example, estimated that approximately seven-tenths of dictionary entries in Vietnamese are idiomatic polysyllabic collocations; a not insignificant proportion of these consist of at least one unavailable term in the sense employed here.

Collocations such as *dế dồ* consist of 'unavailable' parts which do not occur in isolation, cf. *per-cuss-ion*, *astro-naut*, *heckel-phone*.² The semantic contribution of unavailable parts can often be surmised from collocations in which these parts also appear; for *dế* and *dồ* some are given in (2):

²The *heckelphone*, invented in 1904, is a rarely used baritone oboe larger than an English horn and smaller than a bassoon.

- | | | | |
|-----|----|------------------------------------|-----------|
| (2) | a. | <i>đế quốc</i>
emperor nation | 'empire' |
| | b. | <i>hoàng đế</i>
emperor emperor | 'emperor' |
| | c. | <i>đô thị</i>
metropolis city | 'city' |
| | d. | <i>thủ đô</i>
? metropolis | 'capital' |

The four collocations above each consist of unavailable parts, although in all but the last case Nguyễn Đình Hoà (1991) provides a meaning for both parts. Even so, the resultant forms have idiomatic readings regardless of the semantic contributions provided by the unavailable elements within them, much like the sets {*per-ceive, con-ceive, de-ceive*}, {*per-ceive, per-tain, per-cussion*} or {*heckel-phone, saxo-phone, sarruso-phone, xylo-phone*}.³

Collocations such as *vườn tược* 'gardens' and *quốc phục* 'national-costume' are analogous to *huckle-berry* and *carpal tunnel syndrome*. All these have idiomatic readings as well as one unavailable term. For instance, the syllable *tược* is unusable in isolation and occurs exclusively with *vườn* 'garden': the semantic contribution of *tược* is very limited, giving perhaps no more than plurality. In *quốc phục* 'national-costume' or a similar example such as *đỏ kè* 'red-scarlet', the second term has a more identifiable semantic contribution, but is nevertheless limited to only this collocation.

Finally, collocations of free forms such *đèn sách* 'lamp-book' = 'to study' (an exocentric NN compound with verbal syntax) or *xoe tơ kết tóc* 'bind silk weave hair' = 'get married' consist of available terms with suppression of the compositional meaning, cf. *hard-nose* or *take the veil*.

2.1. Separable Collocations and Availability

Exempting idiomaticity and availability from consideration, on what basis, then, is *vườn tược* 'gardens' two 'words' while *huckle-berry* is only one? According to the thesis of atomicity, we can determine if *vườn tược* is one word or two by assessing whether any 'sentence-level description or process' can see into *vườn tược*.

Data such as in (3) are thus immediately relevant (I gloss as XX any unavailable term).

³Because the parts are unavailable, we naturally expect there to be ill-formed combinations such as **de-cussion*.

- (3) a. *Tôi lo vườn tược.*
I care.for garden XX. 'I take care of gardens'
- b. *Tôi lo vườn lo tược.*
I care.for garden care XX. 'I take care of gardens'
- c. *Tôi lo vườn với tược.*
I care.for garden and XX. 'I take care of gardens'
- d. ?? *Tôi lo vườn với lo tược.*
I care.for garden and care.for XX. 'I take care of gardens'

As shown in (3a, b), certain idiomatic collocations are syntactically separable. In this instance, the two terms of *vườn tược* are each preceded by *lo*, 'take care of.' Because suppression of conjunctions is possible in Vietnamese, it might be surmised that (3a) reflects a sentence like (3c) with an overt conjunction linking the two halves of the idiomatic collocation. But native speakers report that (3d), with syntactic splitting of the idiom *and* an overt conjunction, is significantly degraded.⁴ I will thus take it as a working hypothesis that the structure underlying (3a) does not have a null conjunction syntactically.

A variety of contentful predicates can separate idiomatic collocations:

- (4) a. *Tôi xây nhà cửa.* → *Tôi xây nhà xây cửa.*
I build house door → I build house build door.
'I build a house.'
- b. *Tôi không muốn đèn (không muốn) sách.*
I NEG want lamp (NEG want) book
'I do not want to study'

Here the collocations *nhà cửa* 'house-door' = 'habitable, furnished house, i.e. a home' and *đèn sách* 'lamp-book = study' are divisible. The property of being the theme-object of the predicates 'build' or 'not want' is not an inflectional category, and hence cannot be part of that restricted putative 'shared vocabulary' of morphosyntactic properties visible to both word-internal and word-external operations. As Nhân (1976:35) remarks, '... the fact that syntactic rules operate across the morphological-syntactic boundary seems first to challenge the traditional notion of the word ... and secondly, to suggest

⁴However, wh-extraction of one term is impossible for both separable compounds and for phrases with an overt conjunction, so the two types both obey the Coordinate Structure Constraint. Thanks to an anonymous reviewer for pointing out the relevance of this data; thanks to Sonny Vu for soliciting judgments from speakers.

that the boundary between morphology and syntax is indistinct.' To maintain the thesis of atomicity we must claim that *nhà cửa*, *vườn tược*, and *đèn sách* consist of two words, despite the fact that these collocations have either an idiomatic reading or contain unavailable parts.

Not all polysyllabic collocations are divisible in this manner. For example, *xà-phòng* 'soap' (from French *savon*) and total reduplicatives such as *ba-ba* 'tortoise' are not divisible:

- (5) a. **Tôi uống xà uống phòng.*
 I drink sa- drink -von. 'I drink soap'
 b. **Tôi có ba có ba.*
 I have ba have ba. 'I have the tortoise'

Similarly unsplitable are *ô tô* 'car' (from French *auto*) or *cào-cào* 'grasshopper.'

In both these cases the operative restriction is that neither term is 'available' in the sense described above. Although the second term of *vườn tược* 'gardens' is also not available, the restriction on separability appears to require only that at least one term be at least somewhat available. The precise semantic contribution of *tược* is underdetermined, much like *wood-* in *wood-chuck* or *musk-* in *musk-rat* (both folk etymologies from borrowings from Algonquian). But as Aronoff (1976) so succinctly put it, 'what is essential about a morpheme ... [is] ... not that it mean, but rather merely that we be able to recognize it.' Put somewhat differently, once a phonetic string has a 'life of its own'—semantic or otherwise—it may, although need not, become a morphological constituent in its own right.

To summarize, both *đèn sách* 'lamp-book = study' and *vườn tược* 'gardens' are syntactically separable since both contain at least one sufficiently available term. Only one available term is necessary, since the other term acquires autonomy in virtue of being the residue left over when the available term is removed. In the case of *xà-phòng* 'soap' and *ba-ba* 'tortoise', neither term is available and no splitting is possible. In consequence, polysyllabic place names such as *Hà-Nội* 'Hanoi' or *Sai-Gòn* 'Saigon' are completely unsplitable. The task which we turn to in sections 2.2–2.3, then, is a formal characterization of the formal nature of and constraints on the splitting phenomenon.

2.2. Compounds Which are Inseparable

Many idiomatic collocations with fully available terms are also not separable. Collocations which specifically do *not* have an additive or 'dvandva' reading cannot be split into co-ordinate structures. Consider the following:

- (6) a. *Tôi đã ngã (*đã) lòng.*
I PAST fall PAST heart.
'I despaired'
- b. *Tôi sơn nhà (*sơn) tắm.*
I paint house (*paint) bathe.
'I paint the bathroom.'
- c. *Tôi có ống (*có) khói.*
I have tube (*have) smoke.
'I have a chimney.'

(6a) and (6b) cannot be potential co-ordinate structures for syntactic reasons. The first, *ngã lòng* 'fall heart' = 'to despair' is a verbal VN collocation, while the second, *nhà tắm* 'house bathe' = 'bathroom' is a nominal NV collocation.

(6c) is more interesting, being an idiomatic NN collocation which is however not potentially subject to a dvandva interpretation. In *ống khói* 'tube-smoke' = 'chimney' the relationship between the parts is not co-ordinate or additive; rather, a chimney is a tube *for* smoke. Similar 'subordinating' collocations such as *nước mắt* 'water-eye' = 'tear' (water *from* eye) or *nước lửa* 'mountain-fire' = 'volcano' (mountain *of* fire) are inseparable. The hypothetical split form is syntactically well-formed, with a coordinate [V₁ N₁ (and) V₁ N₂] structure, allowing a literal reading such as 'I have a tube and have smoke,' but the idiomatic reading does not survive the split.

The facts illustrated in (6) may serve to establish two things. First, it might be concluded that the unseparability of *ống khói* 'tube-smoke' is evidence that this collocation is a (compound) 'word' and so is opaque to syntactic processes, much like English *bagpipe*. On this view, the separability of *nhà cửa* 'house-door' = 'habitable home' is evidence that this collocation is an idiomatic phrase, much like *French horn* (cf. *He plays the French and English horns*). But this analysis is unsatisfactory insofar as it fails to explain why in Vietnamese only those collocations which require a subordinating structure semantically are also unsplittable.⁵

⁵Put differently, only collocations which do not involve a relation of semantic subordination (argument-taker to argument) can be separated. Heidi Harley points out that this fact may be part of a more general fact: coordinability is permitted only to the extent that the subparts have a similar syntactic and interpretive role:

- (i) Chris turned the oxygen on and the acetylene off.
(ii) ?Kim threw the towel in and the garbage out.
(iii) *The CIA kept a file and tabs on Jane Fonda.

2.2.1. The Encyclopedia in Distributed Morphology

Instead, I will attempt to show that a more satisfactory explanation is available if both separable dvandva collocations and inseparable subordinating collocations are analyzed as consisting of two syntactic terms.

To this end, it will be useful to review certain recent proposals of Marantz (1995, 1997) regarding the interpretation of idioms. On Marantz's proposal, a speaker's knowledge of language must contain a list of idioms (such as were defined earlier) with information regarding their conventionalized meanings: this list is known as the Encyclopedia. The Encyclopedia bears a certain resemblance to the *applicateur d'idiocyncrasie* of Corbin (1987), whose function is to 'apply' conventionalized meanings to words such as *transmission*. But the Encyclopedia differs from Corbin's *applicateur* in several key respects.

First, the Encyclopedia supplies all conventionalized meanings, including those of indivisible constituents such as *atlas* as well as those of phrases like *take the veil*. Second, the meanings so provided must, by hypothesis, be consistent with the meaning imposed by the structure of the idiom in question.

The notion of 'structural' meaning is difficult to state precisely, but has figured prominently in several research programs, including Construction Grammar (Fillmore and Kay 1993, Goldberg 1995) and studies relating to the acquisition of argument structure (Gleitman 1990, Gleitman et. al. 1996, Lidz 1998). In *Distributed Morphology*, it is assumed that syntactic structures are abstract representations without phonetic content (Halle & Marantz 1993). As such they consist solely of categories made available by universal grammar arranged in structures also made available by universal grammar. Following Hale & Keyser (1993), it is proposed that certain of these structural configurations have a canonical 'meaning,' particularly as regards verbal aspect, although the term 'meaning' here must be understood delicately. It is clearly not the case that the entire 'meaning' of a sentence such as *The atlas is on the table* arises from the syntactic structure of the sentence. On the view assumed here however, UG provides a set of configurations of categories and a canonical interpretation of these: all additional meaning is 'encyclopedic'—that is, culturally specific or 'private.' The Encyclopedia provides what might be termed 'semantic detail' beyond what is constructed from universal syntactic-semantic primitives.⁶

While (i) is fine, with *oxygen on* and *acetylene off* being both semantically and syntactically parallel, (ii) and even more so (iii) are degraded, having a joke or zeugmatic interpretation.

⁶Note that *Distributed Morphology*, unlike *Construction Grammar*, holds that structure/meaning correspondences are always universal, never language-specific.

Insertion of phonetic material does not occur until after syntax: the choice of phonetic forms affects only 'conventionalized' meaning or 'semantic detail,' and has no consequences for those aspects of meaning which depend on the choice of configurations and abstract syntactic categories.

Consider now the two types of structures which a collocation of two terms may appear in. In (7a), the two terms are in a co-ordinate structure: neither term is the head. In (7b) on the other hand, one term—in this case the term on the right—is the head, and it projects itself as the category of the collocation:



(7b) is the typical endocentric structure, where the whole is a projection of a part. Some semantic relation must hold in consequence of the inequality in the relationship of α and β here: in a theory with thematic role primitives, one could construe the relationship as 'theta-assignment,' although this interpretation of the relationship is not in fact crucial to the present account. (7a) differs from a standard exocentric structure in that γ , the category of the whole, is the same as the category of both α and β , but γ is not a projection of either α or β . While exocentric structures like (7b) are normally not admitted in X-bar theory (Stowell 1991, Chomsky 1995), I will assume that a structure like (7b) is possible to capture asyndetic (conjunction-less) conjunctions or co-ordinate compounds. A different semantic relation will hold in this instance; whatever the relation in (7a) is, it cannot be the same as what would normally be called a theta-role assignment relation from one part to the other, which requires structure (7b). Instead, (7b) always has an "additive" or "conjunctive" interpretation, where the categories of both α and β have the same syntactic distribution as the category of γ .

2.2.2. Separable and Inseparable Structures

As discussed above, in Vietnamese idiomatic interpretations are preserved in co-ordinate structure only if the conventionalized 'meaning' of the idiom does not require the subordinate structure (7b). For example, although *ống khói* is an idiom with a meaning more specific than 'smoke-tube' it is still not the case that the idiomatic meaning cancels the structural relations holding among the parts of the idiom. Instead, 'tube-smoke' must have the head-modifier relation in syntax (7b), and the Encyclopedia supplements the meaning with such properties as distinguish chimneys from mere smoke-tubes. A fundamental tenet of the proposal is that conventionalized meanings are in-

herently situation-dependent and matters of cultural or personal idiosyncrasy. For each speaker of Vietnamese a chimney is distinct from a mere tube for smoke in ways which are of little interest to the theory of grammar.

Consider now the co-ordinate compounds discussed so far here:

- (8) a. *nhà cửa* 'house-door' = home
 b. *vườn tược* 'garden - XX' = 'gardens'
 c. *đèn sách* 'lamp-book' = study

The Encyclopedia has the capacity to suppress or supply a conventionalized meaning of the minimal constituents of a structure. In the case of (8a), a hypothesis consistent with a dvandva interpretation of *nhà cửa* 'home' is that the Encyclopedia suppresses the conventionalized meaning of the second term *cửa* 'door' and supplies a meaning approximately like 'such amenities as make a house habitable, e.g. furniture.' (This approximates the paraphrase of my consultant.) However, a dvandva or additive interpretation is not really necessary on the present theory: all that matters is that no *subordinating* relation need hold among the parts. Thus, *nhà cửa* = 'habitable home' is consistent with structure (7a) inasmuch as neither *nhà* nor *cửa* is the head of the structure.

Similarly, it is possible that the Encyclopedia supplies to the second term of *vườn tược* 'gardens' the meaning 'garden'. Note that we obtain the result that *tược* by itself is meaningless because the Encyclopedia supplies the meaning of 'garden' to *tược* only when *tược* appears in a co-ordinate structure with *vườn*. Nevertheless it is perhaps imprecise to speak of *tược* as having the meaning 'garden.' It is equally consistent with the present account to assert that *tược* alone has no 'meaning' at all. Again what is crucially important is that whatever *tược* may 'mean', there need not be a syntactic relation of subordination between it and *vườn* 'garden.'

The case of *đèn sách* 'lamp-book' is especially interesting. Neither subpart of this collocation is a verb, yet the whole is syntactically verbal. What matters for separability is the structure must be *headless*: exocentric forms such as *đèn sách* are in fact separable, as predicted (see 4b), although the resulting structure treats the subpart nouns as verbs.

Consider again now unsplitable forms such as *ống khói* 'tube-smoke' = chimney. As we have seen, this collocation *is* splittable, but only with loss of idiomatic interpretation. Hence the issue is not syntax but rather the structures in which idiomaticity survives. Recall now that the Encyclopedia supplies conventionalized meanings to phonological representations within specific syntactic structures. The idiomatic interpretation of 'tube-smoke' = chimney is available only to a [NNN] structure in which one noun is the head and the other is subordinated to the head, i.e. to a head-modifier collocation. Splitting 'tube-smoke' into a series of co-ordinate VPs prevents the

Encyclopedia from supplying the idiomatic reading. For co-ordinate structure idioms, however, the Encyclopedia can supply the idiomatic reading *as long as the structure remains co-ordinate*.

2.2.3. The Syntactic Domain of Splitting

The extent to which idiomaticity is preserved under co-ordination is quite extensive. Consider a VV collocation such as *canh giũ* 'watch-keep' = 'guard':

- (9) a. *Ông sẽ không canh (không) giũ bà.*
 He FUT NEG watch (NEG) keep her.
 'He will not guard her.'
- b. *?*Ông sẽ không canh sẽ (không) giũ bà.*
 He FUT NEG watch (FUT (NEG)) keep her.

In (9a) it is shown that co-ordination of [V] or of [NEG V] preserves idiomaticity. (It remains unclear to me whether the object DP *bà* 'her' is gapped or not in the first co-ordinate.) (9b), however, with co-ordination of [T (NEG) V] is considered ungrammatical by most speakers.⁷

Remarkably, idiomatic VV collocations can interdigitate idiomatic NN collocations, giving a [VNVN VP] structure or a [NVNV NP]⁸:

- (10) a. *Ông sẽ canh vườn giũ tước.* [VNVN VP]
 He FUT watch garden keep XX.
 'He will guard the gardens.'
- b. *Tôi sẽ uống cà nguội phê lạnh.* [NVNV NP]
 I FUT drink cof- cool -fee cold.
 'I will drink the cold coffee.'

Syntactic separation is often used for stylistic effect and colors the expression with the speaker's attitude in a complex way. For example (10b) will be appropriate if the coffee is cold and 'no one wants it': hence the splitting operation in this instance communicates the speaker's disdainful attitude toward the coffee.

⁷Sonny Vu informs me that repetition of tense/aspect markers is ungrammatical according to speakers he has consulted. My consultant however did accept (9b), although certain analogous sentences were judged "wordy."

⁸V denotes either a 'verb' or an 'adjective': adjectives in Vietnamese are syntactically simply (stative) verbs.

In the following example, the VV idiom *dạy dỗ* 'teach-cajole' = 'educate' is interdigitated with the NN idiom *anh em* 'older brother-younger brother' = 'brothers':

- (11) *Ông đã không dạy anh (không) dỗ em.*
 He PAST NEG teach older.brother (NEG) cajole younger.brother
 'He did not educate the brothers'

The literal meaning would imply that only the younger brothers were not cajoled, and only the older brothers were not taught. But the idiomatic or collective reading is also available, according to which it is simply the case that all the brothers were not educated.

Similar to this is (12), in which it is shown that [Adv V] can also coordinate without loss of idiomaticity; here the idiom is *ăn thua* 'win-lose' = 'rival one another':

- (12) *Bà và ông mãi ăn mãi thua.*
 She and he always win always lose.
 'She and he are always rivaling each other.'

Since winning and losing are mutually contradictory, it is clear that 'always' must modify the idiomatic reading of 'rivaling,' even though *mãi* 'always' is repeated in each conjunct.

Idiomaticity is preserved in even larger co-ordinations, such as complex predicates formed with *có thể* 'have ability to':

- (13) *Bà đã có thể dạy đã có thể dỗ con gái.*
 She PAST have ability teach PAST have ability cajole daughter.
 'She was able to educate (her) daughter.'

Although 'wordy' the above sentence is fully grammatical.

There are however strict limits on which co-ordinate structures preserve idiomaticity. Specifically, repetition of the subject cancels the idiomatic reading. Compare (14a) with the idiomatic reading with (14b) without the idiomatic reading:

- (14) a. *Ông sẽ không canh (bà) (?sẽ) không giữ bà.*
 He FUT NEG watch her (FUT) NEG keep her.
 'He will not guard her'
- b. *Ông sẽ không canh bà ông sẽ không giữ bà.*
 He FUT NEG watch her he FUT NEG keep her.
 'He will not watch her and he will not keep her.'
 ≠ 'He will not guard her.'

Whatever the exact difference is between the idiomatic reading 'guard' and the literal reading 'watch and keep,' it is the judgment of my consultant that only the latter is available if the subject (and perhaps also Tense) is repeated. Similarly, compare the following:

- (15) a. *Ông bà dạy cha dỗ mẹ.*
 Grandfather grandmother teach father cajole mother.
 'Grandfather and grandmother educate father and mother.'
- b. *Ông dạy cha bà dỗ mẹ.*
 Grandfather teach father grandmother cajole mother.
 'Grandfather teaches father and grandmother cajoles mother.'

(15a) has the idiomatic and collective reading where the grandparents are educating the parents. (15b) with separation of the coordinate subject into the coordinate predicates has only the literal reading in which the mother is being cajoled (e.g. calmed from crying), but not necessarily instructed.

As confirmation of these semantic judgments, we see that for idioms one of whose parts is not available in the sense defined above, repetition of the subject (16a), or copying of any constituent larger than a clause (16b), induces ungrammaticality:

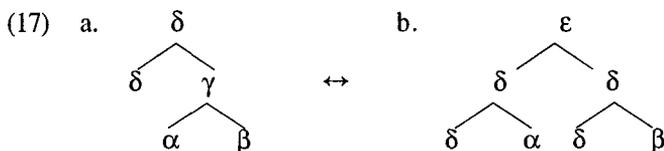
- (16) a. **Ông sẽ không canh vườn ông sẽ không giữ tước.*
 He FUT NEG watch garden FUT NEG keep XX.
 ? 'He will not watch the garden (and) he will not keep the ??'
- b. *Nguyễn đi vườn để học tiếng Pháp*
 Nguyen go garden to study language French
đi tước để học tiếng Pháp.
 go XX to study language French.
 ? 'Nguyen is going to the garden to study French (and) going to the ?? to study French.'

On the account presented here (16a,b) are not so much ungrammatical as they are meaningless, since, I propose, *tước* is supplied with conventional meaning by the Encyclopedia only if it is in a co-ordinate relation with *vườn* 'garden.' As was seen in (9b), repetition of tense/aspect is highly degraded, and repetition of the subject—which naturally entails repetition of an even larger structure—cannot allow an idiomatic interpretation for any speaker. These data confirm a hypothesis advanced by Marantz (1997:208ff.), namely that the structures made available to the Encyclopedia apparently do not ex-

tend beyond the vP or NegP dominating vP.⁹ Consequently the second conjuncts of (16a, b) are not provided with any meaning, although syntactically they may be well-formed, much like Lewis Carroll's famous gibberish poem "Jabberwocky."

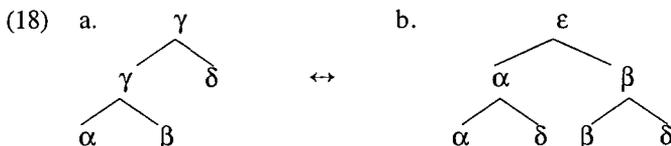
2.3. Formalization

More abstractly, we can model the relation between the non-split and split structures as follows:



A constituent δ taking as complement a co-ordinate structure γ consisting of α , β as parts is equivalent to a coordination of $\delta\alpha$ and $\delta\gamma$. This relationship is recursive; for example (9) exhibits three levels of embedding.

Similarly, a co-ordinate verb can split, attaching its complement to each part:



Here γ , a co-ordinate structure consisting of α , β , takes δ as complement. This structure is equivalent to a co-ordination of $\alpha\delta$ and $\beta\delta$.

The essential insight behind (17) and (18) is that both involve a legitimate exchange of structural relations at PF. In (17) the head-complement relations between A and B (either A or B as head) is exchanged for a co-ordination of head-complement relations between A and the co-ordinate subconstituents of B. In other words, a head-complement relationship between A and B can be distributed over the co-ordinate subconstituents of either A or

⁹As Heidi Harley points out, exactly the same restrictions on idiomatic interpretation arise in the formation of Japanese causatives with *sase* and its allomorphs. As Harley (1995) shows, causative *sase* plus unaccusative predicates may yield idiomatic interpretations, but causative *sase* plus an unergative or transitive (including an already causativized form) permits only non-idiomatic interpretations. For further discussion, see Marantz (1997) and Harley & Noyer (1998).

B. If the subconstituents of A or B are not co-ordinate, distribution is impossible.¹⁰

Added evidence in favor of the relationships above is provided by the phenomenon of conjunction reduction (Nhàn 1984: 340ff). Conjunction reduction relates two idiomatic collocations that share a term with a reduced form in which the shared term occurs only once. Some examples are provided below (Nhàn 1984:343):

- | | | |
|---------|---|---|
| (19) a. | <i>tâm lý sinh lý</i> →
heart reason living reason
'psychological and biological' | <i>tâm sinh lý</i>
heart living reason
'psycho-biological' |
| b. | <i>xuất khẩu nhập khẩu</i> →
exit port enter port
'export and import' | <i>xuất nhập khẩu</i>
exit enter port
'import-export' |
| c. | <i>tiểu công nghiệp thủ công nghiệp</i> →
small work task hand work task
'small industry and handicrafts' | <i>tiểu thủ công nghiệp</i>
small hand work task
'combined small industry
and handicrafts' |

What is important about these forms is that they establish that the relations depicted in (17) and (18) are essentially bi-directional. Idiomaticity is preserved in both cases as long as the structure is co-ordinate.

These cases are no different than more familiar examples of the co-ordination of affixes or stems:

- | | | |
|---------|--|-----------------------------|
| (20) a. | <i>An anti-flea and -lice lotion</i> | (Miller 1992:157) |
| b. | <i>à la cinq- ou sixième entrevue</i>
at the five or six-th interview | (Stendhal, Miller 1992:138) |
| c. | <i>The meat- and potato-eating Scotsman</i> | (Fabb 1984) |

2.4. Discussion

The term 'lexicalized' in the sense of 'having an arbitrary form or an arbitrary meaning' has come to be nearly synonymous with 'object produced by an autonomous Lexicon' with the products of the Lexicon then being syntactically opaque domains, by the thesis of atomicity. What the Vietnamese data so far show is that syntactic objects need not be independently meaningful or available; rather, syntax can manipulate objects which are, from the semantic

¹⁰In the same manner, arithmetic expressions $a*(b+c) = (a*b)+(a*c)$, but $a*(b*c)$ does not necessarily equal $(a*b)+(a*c)$.

perspective, no different from the subparts of such English words as *saxophone*, *huckle-berry*, *musk-rat* or *astro-naut*. We must conclude that arguments to the effect that these latter forms cannot be syntactically constructed—because they contain unavailable terms or have an idiomatic interpretation—are groundless. More generally, arguments that a certain structure is not syntax can only be advanced in the context of a theory which states precisely what syntax can or cannot do in a strictly formal sense.

3. Reduplicatives

So far the discussion has focused on idiomaticity and availability, but an *additional potential criterion for wordhood* is that word may be the domain for conditioning allomorphic choice or morphophonology (i.e. non-automatic phonology). Again, Vietnamese provides a challenge to this attempt at pinning down 'word', since allomorphic and morphological operations apply between splittable co-ordinate structures.

A great many morphemes in Vietnamese have a reduplicative counterpart. As discussed by Văn Lý (1948), Thompson (1965), Phong (1976: 42 ff.) reduplication can be total, or the reduplicant can—in present-day terms—be specified for an onset, for tone, for nucleus, for a rime-plus-tone, or for an onset plus tone:

(21)	Reduplicant specification	structure	example	gloss
a.	None	?	<i>ba ba</i>	'tortoise'
b.	Onset	RED-BASE BASE-RED	<i>bối rối</i> <i>tham lam</i>	'embarrassed' 'eager'
c.	Tone	RED-BASE BASE-RED	<i>đỏ đỏ</i> <i>xổ xốp</i>	'reddish' 'very spongy'
d.	Nucleus	BASE-RED RED-BASE	<i>mập mạp</i> <i>nhút nhát</i>	'fat' 'timid'
e.	Rime & Tone	BASE-RED	<i>nhỏ nhạt</i>	'trivial'
f.	Onset & Tone	BASE-RED	<i>khét lẹt</i>	'strongly burnt'

Moreover, the Base of a reduplicative can be either available or unavailable. For example, the reduplicative *thình lình* 'suddenly' consists of two unavailable parts; in such case it is not immediately obvious which term is the base and which the reduplicant.

For most reduplicative structures, the base is collocated with a particular allomorph of the reduplicant, that is to say, a reduplicant allomorph which has some pre-specified structure:

(22)	Base	Reduplicant	Combination
a.	<i>đỏ</i> 'red'	Ton égal <i>đỏ đỏ</i>	
b.	<i>xấu</i> 'ugly'	Rime = <i>a</i>	<i>xấu xa</i>
c.	<i>bạn</i> 'friend'	Rime = <i>è</i>	<i>bạn bè</i>

In *đỏ đỏ* (22a), for example, the reduplicant allomorph chosen is a prefix specified only for the corresponding 'ton égal'. In the case of *đỏ* the corresponding tone is the high level tone (unmarked in the orthography).

From the point of view of availability and idiomaticity, Vietnamese reduplicatives have analogous forms in English as shown below (data from Marchand 1960):

(23)	Reduplicant specification	Structure	Example
a.	None	?	<i>goody-goody</i>
b.	Nucleus	RED-BASE BASE-RED ?	<i>chit-chat, criss-cross</i> <i>jingle-jangle</i> <i>spick-and-span</i>
c.	Rime	RED-BASE BASE-RED ?	lacking <i>super-duper, teeny-weeny</i> <i>hanky-panky, willy-nilly, hum-drum</i>

The *goody-goody* type includes total reduplicatives with zero available terms. The *chit-chat* and *jingle-jangle* type are nucleus or 'ablauting' reduplicatives with one available term but the *spick-and-span* type has no available term. Finally forms like *super-duper* and *hanky-panky* show rime reduplication with one and zero available terms.

Although morphophonologically the relation between the terms of a reduplicative is unequal (one term is the base while the other is not), from a structural perspective, reduplicatives are co-ordinate forms, since there is no subordinating relation between them in semantic terms. Consider *super-duper* or *hanky-panky* in English. Semantically there is no reason to suppose that *duper* is subordinate to *super* or vice-versa, since *duper* means nothing without *super*.

However, at least in the Northern Dialect, the reduplicative allomorph *-iêc* is highly available with a constant interpretation of 'and such like things/properties/actions,' much like colloquial American English 'and stuff':

- (26) a. *học hiêc* 'study + REDUP(-*iêc*)' → 'study and stuff'
 (Southern dialect: *học hành*)
 b. *bạn biêc* 'friend + + REDUP(-*iêc*)' → 'friends and stuff'
 (Southern dialect: *bạn bè*)

There are no restrictions on the use of this affix: even polysyllabic borrowing like *pê ni xi lin* 'penicillin' can be bases for *-iêc*:

- (27) *pê ni xi lin pê ni xi liêc* 'penicillin and stuff'

Even more conclusive in establishing an independent process of reduplication is that reduplication can take as its base another reduplicative (Phong 1976: 44 ff.):

- (28) a. (*phục* + REDUP(Nucleus = *i*)) + REDUP (high rising tone)
 → *phục phịch* + REDUP (high rising tone)
 → *phục phịch phúc phích*
 'very fat'
 b. (*lừ* + REDUP (Onset = *đ*)) + REDUP (low rising tone)
 → (*lừ đừ*) + REDUP (low falling tone)
 → *lừ đừ lử đử*
 or:
 REDUP (low rising tone) + (*lừ* + REDUP (Onset = *đ*))
 → REDUP (low falling tone) + (*lừ đừ*)
 → *lử đử lử đử*
 'very fat'

The examples in (24) show what Phong (1976) calls 'redoublement en bloc' and Nhân (1984) calls 'top-most expansion'. In this case, the terms of the base are repeated as a constituent, with some modification of one or both terms:

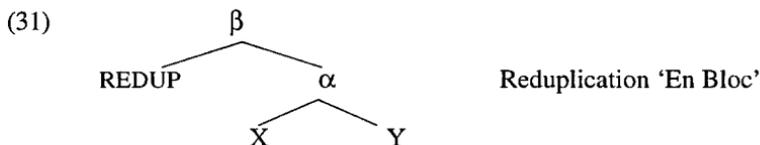
- (29) En bloc: $X Y \rightarrow X' Y' + X Y$ or $X Y + X' Y'$

This type of reduplication nearly always has an intensive reading and the reduplicant can appear prefixed or suffixed. (28b) gives an example where either is acceptable.

There are several subvarieties of reduplication en bloc, depending on the change induced in the reduplicant. The forms in (28) all have a reduplicant specified only for tone. Where the reduplicant is specified for a rime *R* then either term *Y* or both *X* and *Y* can have *R* overwritten:

- (30) a. *liú lo* + REDUP(Rime = *ư ò ng*) → *liú lo liú lươ ng*
 'chirp incessantly'
 b. *bông lông* + REDUP(Rime = *ang*) → *bông lông bang lang*
 'wander aimlessly' (Nhàn 1984: 252)

Formally a reduplication en bloc conforms to the same structural requirements as were discussed in section 2.3 for syntactic reduplications:



Here the constituent REDUP takes α as its complement (and target). Phonetic realization maps this configuration to a string in which one or both constituents *X* and *Y* are overwritten by the reduplicative affix:

- (32) [RED [α X Y]] → RED(X) RED(Y) X Y

A second type of reduplication is termed 'redoublement intercalé' by Phong and 'atomic-expansion' by Nhàn:

- (33) *nhút nhát* → *nhút nhút nhát nhát* 'timid'
hớn hớ → *hớn hớn hớ hớ* 'cheerful'

In intercalated reduplication, the first term is repeated twice and then the second term twice:

- (34) Intercalé: *X Y* → *X' X Y Y'* or *X X' Y Y'*

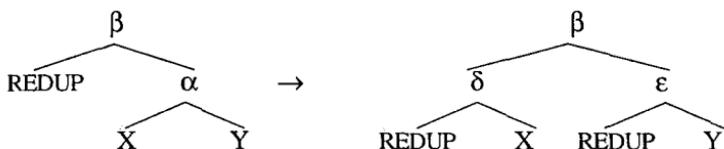
The exact meaning ascribed to the intercalated reduplicatives varies from author to author. Phong (1976) calls it 'attenuative' whereas Nhàn (1984) translates with 'consistently,' 'repeatedly,' or 'excessively.'

Some examples of total, intercalated reduplication are shown in (35). In (36) reduplication overwrites the base with tone only, and in (37) with onset only:

- (35) *nhút nhát* → *nhút nhút nhát nhát* 'timid'
hớn hớ → *hớn hớn hớ hớ* 'cheerful'
- (36) a. REDUP (ton égal) + (*lỏng lẻo*) → *long lỏng leo lẻo*
 'loose'
 b. REDUP (ton égal) + *ngộ nghĩnh* → *ngồ ngộ nghĩnh nghĩnh*
 'beautiful' (Phong 1976:51)
- (37) a. REDUP (Onset = /l/) + *khệnh khạng* → *lệnh khệnh lạng khạng*
 'walking in an air of exceeding importance'
 b. *xơ xác* + REDUP (Onset = /r/) → *xơ rơ xác rác*
 'frayed' → 'ragged'

Again, the intercalated reduplications conform to the formal properties of syntactic reduplicatives discussed in section 2.3. Specifically, intercalation results from the distribution of REDUP as a sister of its complement α to a sister of both co-ordinate daughters of α :

- (38) Reduplication 'Intercalé':



On this basis, I conclude that reduplicatives have the same expansion pattern as other *dvandvas*, where by '*dvandva*' I mean simply any co-ordinate structure with no internal thematic relation. The reduplicants are both morphemes in the morphophonological sense and also syntactic atoms.

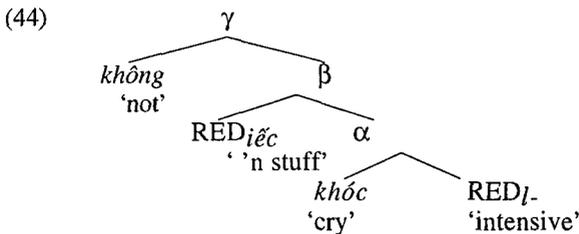
Not surprisingly, reduplications of reduplications are subject to further expansion. Examples (39–43) show separability by *hơi* 'somewhat', *không* 'NEG', and *sẽ* 'FUT.'

- (39) *Cô ấy hơi nhút (*hơi) nhút (hơi) nhát (*hơi) nhát*
 RED'-RED'-RED-timid = a little timid
 'She is sort of a little timid.'
- (40) *Hàng cây hơi ần (*hơi) ần (hơi) hiện (*hơi) hiện*
 RED-appear-RED-disappear = shimmering
 'The row of trees is sort of a little shimmering.'

- (41) *Em bé không khóc (*không) lúc (không) khóc (*không) liếc*
cry-RED-cry-RED' = crying and carrying on
'Baby brother is not crying and stuff.'
- (42) *Cai nơ không lỏng (*không) lỏng (không) leo (*không) lẻo.*
loose-RED-RED'-RED'' = extremely loose
'The knot is not extremely loose (but still loose).'
- (43) *Ăn mày sẽ lông (*sẽ) bông (sẽ) lang (*sẽ) bang*
vaga-bond-RED-RED = 'wander aimlessly'
'The beggar will wander aimlessly.'

As can be seen above, a form like *nhút nhút nhát nhát* 'timid' can be separated only once by a c-commanding element such as *không* 'NEG.' The judgment of my consultant suggests that this holds regardless of the form (intercalated vs. en bloc) of the reduplicative, and regardless of other factors such as the whether the reduplicant is a prefix or suffix, or total or partial.

To explain these judgments we have only to invoke the notion of cyclic application. Consider the derivation of ill-formed **không khóc không lúc không khóc không liếc* 'not crying and stuff'. The underlying constituency is given in (44) and the derivation is shown in (45):



- (45) [_γkhông [_β REDiếc [_α khóc+REDl-]]
 [_γkhông [_β REDiếc [_α khóc+lóc]] α-cycle
 [_γkhông [_β khóc+lóc khóc+liếc]] β-cycle
 [không khóc+lóc] [không khóc+liếc] γ-cycle expansion

Supposing the the expansion rule is cyclic, it will in this instance apply on the cycle defined by the constituent γ in (44), giving a single expansion into well-formed [*không khóc+lóc*] [*không khóc+liếc*] 'not crying and stuff.' But further expansion will be impossible, because doing so will require *không* 'NEG' to 'see into' a subconstituent fully contained on the preceding cycle. This provides strong evidence that the structural configurations depicted in (45) are correctly viewed as syntactic, part of a generalized morpho-syntax in which the notion 'word' plays no role.

4. Summary

On the basis of such well-known examples as *per-ceive*, *con-ceive*, *de-ceive*, Aronoff (1976) argued persuasively that morphemes do not require any fixed meaning nor do they require productive combinatory possibilities. They are purely formal elements. Abandoning the criteria of idiomaticity and productivity as criterial has extremely far-reaching consequences. On purely formal grounds we have found no evidence for separating 'word'-sized units from 'morpheme'-sized units in Vietnamese. Headless structures of all types are syntactically separable, whether dvandva, exocentric, reduplicatives, or reduplicatives of reduplicatives:

(46)	<i>nhà cửa</i>	'house-door' = home
	<i>vườn tược</i>	'garden-XX' = gardens
	<i>đèn sách</i>	'lamp-book' = study
	<i>nhút nhát</i>	'timid-RED' = timid
	<i>nhút nhút nhát nhát</i>	'timid-RED-RED' = rather timid

All these must be treated as syntactic constructions. Yet many have idiomatic reading and many contain unavailable parts, both typically construed as indicating 'lexical' status. Moreover, the relation of arbitrary allomorphic selection obtains between elements which are, on these grounds, syntactic atoms, and the morphophonological process of reduplication takes as its target elements which are syntactic atoms.

In sum, the criteria normally used to distinguish 'word' structures from 'phrase' structures have no force in Vietnamese, showing that the language learner cannot rely on any of these to distinguish a putative module of morphology from the syntax. Instead, principles of a generalized morphosyntax play an important role in determining the well-formedness of reduplicative and co-ordinate structures. An essential problem for future work is therefore whether such an enriched theory of morphosyntax can, for all languages, fully replace the set of operations normally imputed to the Lexicon.

References

- Aronoff, Mark (1976). *Word Formation in Generative Grammar*. MIT Press, Cambridge.
- Chomsky, Noam (1957). *Syntactic Structures*. Mouton, The Hague.
- Chomsky, Noam. (1995). *The Minimalist Program*. MIT Press, Cambridge.
- Corbin, Danielle (1987). *Morphologie dérivationnelle et structuration du lexique*. Niemeyer, Tübingen.

- Di Sciullo, Anna-Maria & Edwin Williams (1987). *On the Definition of Word*. MIT Press, Cambridge.
- Fabb, Nigel (1984). *Syntactic Affixation*. PhD dissertation, MIT.
- Fillmore, Charles J. and Paul Kay (1993). *Construction Grammar*. Manuscript, University of California, Berkeley.
- Gleitman, Lila R. (1990). "Structural Sources of Verb Learning." *Language Acquisition* 1:1-63.
- Gleitman, Lila R., Henry Gleitman, Carol Miller and Ruth Ostrin (1996). "Similar, and similar concepts." *Cognition* 58:321-376.
- Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. The University of Chicago Press, Chicago.
- Halle, Morris & Alec Marantz (1993). "Distributed Morphology and the Pieces of Inflection," in K. Hale and S.J. Keyser, eds., *The View From Building 20: Essays in Honor of Sylvain Bromberger*. MIT Press, Cambridge, 111-176.
- Harley, Heidi. (1995). *Subjects, Events and Licensing*. MIT PhD dissertation.
- Harley, Heidi and Rolf Noyer (1998). "Licensing in the Non-Lexicalist Lexicon: Nominalizations, Vocabulary Items and the Encyclopaedia," in H. Harley, ed., *Papers from the UPenn/MIT Roundtable on Argument Structure and Aspect: MIT Working Papers in Linguistics* 32, 119-138.
- Hoà, Nguyễn Đình (1991). *Vietnamese-English Dictionary*. C.E. Tuttle, Vermont.
- Lidz, Jeffrey (1998). "Constraints on the Syntactic Bootstrapping Procedure." Manuscript, University of Pennsylvania.
- văn Lý, Lê (1948). *Le parler vietnamien*. Huong Anh, Paris.
- Marantz, Alec (1995). "A Late Note on Late Insertion," in Young-Sun Kim, et al., eds., *Explorations in Generative Grammar*. Hankuk Publishing Co., Seoul, pp. 396-413.
- Marantz, Alec. (1997a). "No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon, in A. Dimitriadis et al., eds., *Penn Working Papers in Linguistics* 4: 201-225.
- Marchand, Hans (1960). *The Categories and Types of Present-day English Word-formation*. Otto Harrassowitz, Wiesbaden.
- Miller, Philip (1992). *Clitics and Constituents in Phrase Structure Grammar*. Garland, New York.
- Nhàn, Ngô Thanh (1984). *The Syllabeme and Patterns of Word Formation in Vietnamese*. PhD dissertation, New York University.
- Noyer, Rolf (1997). *Features, Positions and Affixes in Autonomous Morphological Structure*. Garland, New York.
- Phong, Nguyễn Phú (1976). *Le syntagme verbal in vietnamien*. Mouton, Le Haye.
- Stowell, Timothy (1981). *Origins of Phrase Structure*. MIT PhD dissertation.
- Thompson, Laurence (1963). "The Problem of the Word in Vietnamese." *Word* 19:39-52.

- Thompson, Laurence (1965). *A Vietnamese Grammar*. The University of Washington Press, Seattle.
- Vu, Sonny (1998a). "Prosodic Structure in Reduplication: A Unified Account of Vietnamese Reduplication." Manuscript, MIT.
- Vu, Sonny (1998b). "Splitting Heads: Separability as a Case of Covert Head Movement." Manuscript, MIT.

Department of Linguistics
University of Pennsylvania
Philadelphia, PA 19104-6305
rnoyer@babel.ling.upenn.edu

The *University of Pennsylvania Working Papers in Linguistics (PWPL)* is an occasional series produced by the Penn Linguistics Club, the graduate student organization of the Linguistics Department of the University of Pennsylvania.

Publication in this volume does not preclude submission of papers elsewhere; copyright is retained by the authors of the individual papers.

Volumes of the Working Papers are available for \$15, prepaid. Please see our web site for additional information.

The PWPL Series Editors

Alexis Dimitriadis
Hikyoun Lee
Christine Moisset
Alexander Williams

How to reach the PWPL

U. Penn Working Papers in Linguistics
Department of Linguistics
619 Williams Hall
University of Pennsylvania
Philadelphia, PA 19104-6305

*<http://www.ling.upenn.edu/papers/pwpl.html>
working-papers@ling.upenn.edu*