aux_ing or *aux_en* structures). For convenience, these have been summed and added to the bottom of Tables 4.27 to 4.30 under the rubric of *AUX*. No attempt has been made to decompose these instances of auxiliary be into the various possible forms in the paradigm. It should be clear, however, from the figures provided that be as a copula is much more frequent and widely distributed than be as an auxiliary.

So far, no figures have yet been compiled for the non-realization of the copula in environments where it could be expected to appear, nor has any investigation of a possible hierarchy of contexts favouring the production of be been conducted. Other studies have postulated that copular be emerges first in pre-adjectival environments and subsequently appears in pre-nominal ones [129]. The research could be very usefully extended in the above directions.

Despite the limitations noted above, a number of conclusions emerge quite clearly from the analysis.

4.7.2 'Is'

Is, and its bound variants, is the most widely distributed, and by far the most frequent, form of be. Except in the case of My, where the counter-examples are probably formulaic, every informant who produces be in some form produces tokens of lexical is. While there is insufficient longitudinal data on the acquisition of is, with only two informants not producing the form in the first interview, on the basis of the cross-sectional patterning it is fairly probable that is is the first form of be to be acquired.

There is, however, a very striking discrepancy between the Polish speakers and the Vietnamese speakers in regard to the relative frequency with which is is produced. This discrepancy is evident, possibly to a lesser degree, with other forms of be as well.

Although counts for null copula have not been done, it is clear from a perusal of the transcripts together with the evidence from the relevant tables, that Vietnamese speakers very frequently fail to produce the copula in structures where the target requires it.

Within the theoretical framework of the *ZISA* Project, there is, as described in 2.1.8, a distinction drawn between developmental and variational features [36]. Developmental features are necessarily acquired in a particular order because they depend on increasingly complex and interdependent speech processing prerequisites. Variational features, however, are not constrained in the same way. The copula in German was classified in the *ZISA* project as a variational feature. Once acquired, the conditions affecting its suppliance are different from the computational and cognitive processes that determine, say, word order phenomena.

If this developmental/variational distinction is accepted there is no reason not to assume that the English copula also has the status of a variational

feature. In the absence of such a distinction, one would be left with the problem of explaining why a learner such as Phuc, who has a high *ASLPR* rating and displays considerable mastery of likely developmental features, such as modals, verb morphology and subordination, performs so poorly in this particular area.

Variational features were originally envisaged as being determined by sociolinguistic factors—exposure to native speech, social and psychological distance, etc. These factors may well have a bearing on the low rate of copula realization amongst the Vietnamese speakers.

One further factor, however, is almost certain to have a considerable influence. This is first language: Vietnamese does not require the copula in equative structures involving predicate adjectives. There are two ways of looking at the effects of first language in this case. One is to conclude that the copula, as a variational feature, and one therefore largely unconstrained by semantic or processing factors, is highly vulnerable to influence by extraneous factors, such as first language rules. A second approach, which adheres more closely to principle that variational features reflect sociolinguistic factors is to include such traits as reluctance to abandon first language features amongst the sociolinguistic factors, perhaps under some such heading as social distance. Under both approaches, variational features become likely targets for interference or transfer phenomena. The main difference is that the second approach makes more specific proposals as to when these phenomena would manifest themselves. These proposals should be testable. Amongst the present group of Vietnamese informants both Vinh and Phuc, who come from an educated middle-class family, seem to perform better in regard to copula production than the other informants. This offers tentative support to the second of the above approaches. It would be interesting to check the figures for copula production more closely, and to study this feature in a larger group of speakers to see how well the proposal is borne out.

4.7.3 'I'm' and 'Am':

As can be seen from Tables 4.27 to 4.30, amongst the informants ranked lower on the *ASLPR*, a good many tokens of I'm are demonstrably pronominal. In some cases, this is so for an informant's whole output of this particular form. In all four interview groups this phenomenon extends at least to level 1 on the *ASLPR*, and beyond it in the case of one or two informants.

With regard to am, it is frequently more difficult to find clear evidence of its use as a proform fragment. It appears, however, that the Polish speakers are more successful in manipulating this form than the Vietnamese speakers, whose output includes several cases of I am as a pronoun. This is, of course, in line with the overall pattern for copula insertion for the two language groups. In the cases of My and Duc, there are one or two cases where it

is difficult to decide if I am is a pronoun, or am is somehow being confused with have. This is because of utterances such as the following:

becau(se)—I am no money

- my.2: [844] -

4.7.4 'Are'

The discrepancy between the two language groups in the frequency with which are occurs is not so great. The probable explanation for this lies in differential discourse patterns for the two groups, for which a rough index can be found in the distribution and frequency of pronoun use. Tables 4.76 to 4.80 show that you, and to a lesser extent they, are both more widely distributed and much more frequent for the Vietnamese speakers. (The implications of this are discussed in 4.17). The much larger number of possible occasions for are in the discourse of the Vietnamese speakers probably compensates for the lower likelihood that they will produce this form of the copula.

It should also be noted that the sole tokens of are in **my.1**, **minh.1** and **ka.1** are in formulaic questions, and are therefore dubious. In **hoa.1** and **es.1** two of the three tokens produced are also in formulaic questions; and in **jb.1** the count is three out of four. Although we can assume that the form is established for these informants, this affects the status of the counts, which after adjustment sink to a single token per interview for all informants under 1+ on the *ASLPR*, with the sole exception of Dung.

4.7.5 'Was' and 'Were'

For the Polish speakers, the distribution of was is quite wide, beginning at *ASLPR* level 0+. For the informants above level 1+, there is a marked increase in the number of tokens produced. For the Vietnamese speakers, the distribution is rather more ragged (with the lower threshold still at about 0+ however) and the frequencies are considerably lower. Once again, this is presumably a reflection of the overall discrepancy for the two groups; there may, however, be a greater tendency for the copula to be inserted where a non-present tense has to be indicated, which would have the effect of decreasing the discrepancy to some extent.

For were, the distribution is narrower and more ragged. This is true for both groups, though more so for the Vietnamese. Nevertheless, there exists an implicational relationship between was and were; if the latter is produced, so is the former.

It is also worth noting that distinctions involving person and number, such as that between was and were seem to emerge at different stages for copular and auxiliary uses of be. For a more detailed discussion of this,

and other phenomena involving both be and verbs in general, the reader is referred to 4.2

4.7.6 'Been'

Been is basically found in two structures. The more widely distributed of these is in (largely formulaic) tokens of the present perfect:

> they ... er... ha(ve) been here... two year(s)
>
> - es.1: [154] -

> Yeah, my name is Dung... I have been in Australia, er, for two year(s)... nearly two years
>
> - dung.1: [8–9] -

The other use of been, which obviously results from a simplification of the present perfect structure is as a preterite form of be:

> Die bed... the rooms been... no big... no big
>
> - ka.1: [680] -

> Thailand... we been... er... in refugee camp... Songkhla
>
> - tam.1: [129] -

The distribution of been is fairly ragged, which is what one would expect for an item found almost exclusively in formulas.

A more extended discussion of perfects is to be found in the section on *aux_en* in 4.2

4.7.7 'Be'

The principal significance of the infinitival form be itself, is that since it almost always occurs as the complement of some other verb, or in conjunction with a modal, it can serve as a quite useful index of clausal complexity. Like its finite forms, however, it is subject to variable realization patterns and is both more widely distributed and more frequent amongst the Polish speakers. The variable nature of be has, of course, to be taken into account for it to produce reliable results as an index of complexity; some monitoring of the production of a form such as is would have to be carried out first.

For speakers with a high degree of copula insertion be may emerge as a form at about the same time as modals like will and must begin to be used. The reader is referred to 4.9 for more discussion of this topic.

4.8 Summary

While there are certain regularities in the production of the copula, and evidence for differential treatment of be as a copular form and an auxiliary, it appears to be the case that a good deal of variability will be manifested in the suppliance of the copula, once the processing prerequisites for its production have been met [128]. This variability seems to be partly indexable to learner type and to psycho-sociological factors, and partly a result of first language patterns.

4.8.1 Note on the Tables

The following points should be kept in mind when consulting Tables 4.27 to 4.30:

1. Tokens apparently involving contracted forms of be, such as I'm or what's, may, in some cases, very possibly be monomorphemic noun phrases or proforms. Where this is definitely the case, the counts in question are suffixed with a *P*.

2. It should also be kept in mind that in some cases apparent lexical tokens of be may be artifacts of the transcription and would be more properly described as part of monomorphemic noun phrases or proforms. For example, am might be a fragment of the proform I am, rather than a verb. These phenomena are more likely to characterize the output of the less proficient or more simplifying informants.

4.9 Modals

4.9.1 Distribution of Modals

Modals, because of their own special properties and because of the pedagogic attention they receive have been allotted a section to themselves. The basic data relating to these elements of the verbal system is summarized in Tables 4.31 to 4.34.

As with the other interview group tables, Tables 4.31 to 4.34 are organized according to*ASLPR* rating. In the case of modals, particularly with the Vietnamese speakers, this provides some encouragingly implicational patterns.

It is worth remarking at this point that in general it appears to be the case that the Vietnamese speakers pattern more regularly than the Polish speakers in their linguistic behaviour. It is conceivable that this is due to either:

Informant	is	ks	zj	jb	jr	ka	bb	es	aj	kb	lj	mm
is	11	40	12	20	74	40	156	51	43	85	85	66
it's		5			1	2	3	4	6	18	16	45
he's					1	1		1	1	3	1	
that's		1				1					15	1
NP's			1			2		1		1		
what's		2		2					1			
she's								1		2		2
I's					2							
isn't											25	
I'm	1P	7P	1?	1P	2P		6P	2	4	7	14*	4
am		1		2		7	5		6	2		
was		4	2	6	1	9	1	11	62	45	47	58
wasn't						2					5	1
are		2*		4*		1*		3*	9	3*	4	5
be				2*	1?				7	6	18*	7
were						1		2	2	2		2
weren't						1						
been				1		1*		4		3		
AUX				4*	3*			1	15	7	1	20

Table 4.27: Distribution of Copular Forms: Polish Informants—Interview 1

NOTES:

1. Tokens of "are" occur in formulaic questions in the cases of **ks** (two tokens), **jb** (three tokens), **ka**, **es** (two tokens), and **kb** (three tokens).
2. Tokens of "be" occur in the locution "will be" in the cases of **jb** and **lj** (five tokens).
3. The aux_ing tokens are formulaic in the case of **jb** and dubious in the case of **jr**.
4. With **ka** "been" is a preterite rather than a participle.
5. With **lj** some tokens of "I'm" are definitely proforms.

Informant	is	ks	zj	jr	ka	jb	bb	es	aj	kb	lj	mm
is	34	64	42	66	40	44	39	47	69	74	96	81
it's	4	2	1	9	4	3	87	13	6	30	23	22
he's	2			2		2	1	6	1	5	1	
she's				1	1	2		2		5		2
that's		1	1			1			1		6	
isn't								1			11	1
NP's									1			2
what's		3										
when's							1					
was	4	2	1	4	9	12	13	11	21	23	25	12
wasn't					1							
I'm	2?	1P	1P		1?	3*	9		10	5	6*	3
am					1	1	1	2	2	7		3
are			1	2	1	1	1	6	10	11	9	17
we're								1				
be			3	7		3*	18*	1	15	3	74*	9
been				1	3*			1	2			
were						1	1			3		1
AUX				4	2F	5*		2	11	15	5*	20

Table 4.28: Distribution of Copular Forms: Polish Informants—Interview 2

NOTES:

1. Tokens of "be" occur in the locution "will be" in the cases of **jb**, **bb** (all tokens) and **lj** (thirty-seven tokens).
2. The *aux_ing* tokens are formulaic in the case of **lj** and dubious in the case of **jb**.
3. With **ka** "been" is a preterite rather than a participle.
4. With **lj** some tokens of "I'm" are definitely proforms. In the case of **jb** all tokens of "I'm" are subsequently corrected to "I", indicating a period of transition for this informant.

Informant	Van	My	Duc	Dung	Minh	Sang	Hoa	Vinh	Tam	Canh	Long	Phuc
is			9	2	4	2	6	30	1	13	12	27
that's								4		23	63	29
he's							3	1		1	11	
what's					1		1		1			
it's								2		2		13
NP's							1					2
she's							1	1				
I's											1	
we's											1	
I'm			1P		2	2	29P	7	5	19*	9*	1
am		3?		1	3*	3		1		5		17
are		1*		5*	1*	2*	3*			40	5	31
was			1?	1		2		6		5	1	21
were			1?					1		3		13
been				1	1				1*	6	9*	
be										3	21*	3
AUX										8*	2*	7

Table 4.29: Distribution of Copular Forms: Vietnamese Informants—Interview 1

NOTES:

1. Tokens of "are" occur in formulaic questions in the cases of **My**, **Dung** (two tokens), **Minh** and **Hoa** (two tokens).
2. Tokens of "be" occur in the locution "will be" in the cases of **jb** and **lj** (five tokens).
3. The *aux_ing* and *aux_en* tokens are dubious in the case of **Canh** and **Long**.
4. With **Tam** and **Long** "been" is a preterite rather than a participle.
5. In the cases of **Hoa** (all tokens), **Canh** (two tokens), and **Long** (at least four tokens), tokens of "I'm" are definitely proforms.

Informant	Van	Duc	Sang	Minh	My	Hoa	Dung	Vinh	Tam	Canh	Long	Phuc—
is	1	1		6			1	14	1	10	1	23
it's		3		3		1		1	2	7	1	7
that's					1	1		1		6	14	19
he's				2		1				6		
NP's										2		1
what's							1					
who's							1					
where's								1				
we's									1			
I'm	4P	4P	7		1	38P		5		14*	1	4
am	3*	1*			6*		1*			1		6
be			1			2?		1			5	14
are		1?			1?		1			1		4
was			1				1	1				
been									1*	1		1
AUX											1*	6

Table 4.30: Distribution of Copular Forms: Vietnamese Informants—Interview 2

NOTES:

1. The *aux_ing* tokens are dubious in the case of **Long**.
2. With **Tam** and **Phuc** "been" is a preterite rather than a participle.
3. In the cases of **My** and **Dung** tokens of "am" are fragments of the proform "I am".

1. The fact that they have had, in nearly all cases, considerably longer periods of exposure to English than the Polish speakers, which has led to the "stabilization" of their interim grammars, and/or:

2. The fact that, unlike the Polish speakers, they were not (with the exception of three of the less proficient informants who were attending part-time courses) currently receiving instruction. If this latter consideration is a valid one, then it tends to indicate that instruction is capable of "disturbing" naturalistic learning processes. This disturbance may be explicable in Labovian terms as the result of increased linguistic "self-consciousness" [149] (see also 3.4.2). Whether or not such disturbance is a positive or a negative phenomenon must remain an open question: there is, of course, no *a priori* reason why it should be negative.

While the Vietnamese speakers pattern more regularly, the patterns for the use of modals for the two language groups are basically similar, with one exception. This concerns the relative frequency of must and can. For the Vietnamese speakers overall, can is their highest frequency item, both in gross and distributional terms. For the Polish speakers, this place is occupied by must—see Tables 4.34 to tables 4.32. This overall state of affairs is reflected in the separate counts for the use of these items for the*ASLPR* range from 0+ to 1+. For the Polish speakers above level 1+ this discrepancy in the use of can and must is considerably smaller, as can be seen from the tables.

As far as can be determined, there is no differential pattern in the semantics of these two items as they are used by the two groups which could explain this divergence. The reason for it—it is tempting to suggest—may be in fact cultural. Whorfian treatment of linguistic data, however, can be a dangerous undertaking (consider Lakoff's anecdote about the mysterious language with an extraordinarily large set of verbs for describing modes of destruction—which turned out to be English), and will not be pursued here (cf. the work of Whorf for an extreme example of the "language shapes thought" position[183]).

No very detailed analysis of the semantics of the modals which appear will be provided here. However, it is probably worth noting down some of the salient points concerning each of these modals as they are used by the informants in the study. The observations made here are based on data from the first round of interviews. Some evidence from the second round interviews will also be adduced.

4.9.2 'Can':

In general the semantics of can appear to be fairly standard. It is produced in a variety of syntactic environments from declaratives to questions. In most

cases it is followed by a main verb; occasionally it is used anaphorically:

Yes I can

- dung.1: [78] -

The principal semantic function of can is to indicate capacity. In informants higher up on the*ASLPR* scale can is also used on occasion to indicate permission (**long.1**), and probability (**canh.1**). Not surprisingly, can is sometimes used in past contexts, rather than could. It should be noted that in its function as an indicator of capacity can provides a natural means of expressing propositions about the future.

4.9.3 'Must':

In the majority of cases, must is used, in a more or less standard way, to indicate obligation. It does not seem to be used to express strong probability, however. Only one informant, Long, uses must where another modal is definitely called for:

I hope we... we must go to school one day

- long.1: [374] -

We shall return to the possible implications of this overgeneralization shortly. There are, however, a number of very puzzling tokens involving must, in the output of Long and Tam. These involve the curious locution must have to:

You must have to learn more, Tam... stop, please

- tam.1: [805] -

In all cases, (four in **tam.1**, one in **tam.2**, two in **long.1**), the final element in the structure is learn. In **tam.1**, it transpires, learn only appears in the locution have to learn, and likewise have to, so that it is very probable that have to learn is an idiosyncratic realization of learn. Similarly, in **long.1**, only two or three of the twenty-five tokens of learn are not have to learn, and the same conclusion can therefore be drawn as for Tam. If have to learn is in fact an idiosyncratic form of learn then the behaviour of must is unexceptional in these cases too.

The only notable non-standard semantic feature of must is that it is used in past contexts. Straightforward examples of this would be:

In, um, Songkhla.../... I mus(t), er, sell ci(ga)re(tt)es

- sang.1: [202] -

everybody.../...must go in Army

- long.1: [111–12] -

In their work on the acquisition of German as a second language, Klein and Dittmar found that müssen was "apparently used as a substitute for morphological tense markers of the verb", in particular as a past marker [155]. While tokens of "must" occurring in past tense contexts in the present corpus are by and large similar to the examples given above, and probably ascribable to the simple absence of an appropriate past form for must, there is some possibility that a similar process of overgeneralization may be operating in the case of English must, at least for some informants. Thus, in addition to the use of must in a sentential complement of hope (cited above) in **long.1** there are tokens such as the following:

an(d) we Army before we haven't got any money.../...but we must looking, you know, because we...we think .../...I get fisherman after that I can drive go out

after that I mus(t) thinkin(g), you know, an(d) about not be long...about two, t(h)ree week(s) I know him, you know, an(d) after that understand very well

- long.1: [188–90] & [497–9] -

These tokens (in which the verb is also marked with the past *-ing* marker), suggest that must is possibly essayed as a lexical (past) tense marker in English also. It should be emphasized that this conclusion is quite tentative, and would require further investigation. Dittmar's explanation for the use of müssen as a tense marker in German is morphological. He suggests that the phenomenon "may be explained by the fact that müss has the comfortable property of having the same form in the first and third person singular" [155]. Clearly, this explanation would not hold for English, where modals are never marked for person. An explanation for English (and an alternative explanation for German) can be found in the semantic and systemic properties of must or müssen. This is as follows. First, must is acquired early for functional reasons. Second, there is evidence that lexical morphemes are preferred over bound morphemes—this is a consequence of the learner's tendency to optimize form and function. Third, of the lexical morphemes available for overgeneralization must is the most appropriate, since will is tense-specific and can involves specific propositions about the speaker himself, unlike must, which involves more general propositions about the speaker's environment, and is arguably less marked semantically.

Whatever the status of the above proposal, Dittmar's predicated order of acquisition for modals in German (on the basis of cross-sectional data) is, as we shall see, quite similar to the projected order for English. Dittmar's basic findings are:

> müssen, wöllen, and können are the modals applied with the greatest frequency; (b) wöllen and können are applied at an earlier stage than müssen; (c) mögen and söllen are acquired very late; and (d) müssen contributes more than 50% to the applications of the rule [155].

This suggests that some possibly universal hierarchy of semantic markedness helps determine the acquisition and use of modals. Data from other languages would be interesting to examine in this connection.

Further reference to must will be made in the section on have to.

4.9.4 'Will':

Will is used largely in utterances with a future temporal reference, and in this sense could be said to be fairly standard. On occasion, will appears in utterances whose time reference is future-in-the-past, which clearly represents an overgeneralization of its use (cases of this occur in **minh.1** and **phuc.1**).

> I to learn English, er... about six month(s)... yeah... I will, er, speak Englis(h) easy
>
> - minh.1: [480–1] -

> you know... their relatives... they will be happy to see Vietnam again ... their country lan(d)
>
> - phuc.1: [874–5] -

In several of the informants higher on the ASLPR will occurs in the result clause of conditional structures, (**vinh.1** and **tam.1**), and in the first interview with Phuc (who perhaps exhibits the most development in use of modals of all the informants) will also appears, correctly, in a habitual-predictive context:

> usually they will put you into zhaol
>
> - phuc.1: [469–70] -

However, while will exhibits generally standard, if restricted, semantics, its behaviour from a developmental viewpoint requires comment. Thus, a number of informants who produce will, produce it only in the locution will be. This is particularly common amongst the Polish speakers, but it is also true of Long. The resulting cases are sometimes standard (superficially) and sometimes not, as in these two examples from **lj.1**:

I will be at home three years

- lj.1: [128] -

I will be have ten years

- lj.1: [319] -

Informants whose entire production of will involves will be tokens are Jerzy B., Barbara B. (second interview—no tokens in the first), and Ludwiga J. Besides these, a number of other informants, while producing will alone, also produce significant numbers of tokens of will be. These are Jan R. (two out of five), Andrzej J. (thirteen out of twenty-three), and Krystyna B. (four out of six in Interview One). In the case of Andrzej J., none of the will be tokens are non-standard. In other cases, will be is sometimes used with main verbs, producing non-standard structures.

This morphological peculiarity is explicable in the following terms. First, learners tend to acquire and use formulas in which be appears, probably because common utterances involving will often include be. Examples of such formulas would be:

Wi(ll)... maybe will be better... I don't know

- bb.2: [77] -

I will be looking for... for job

- jb.1: [275] -

not be long de com(p)any will be closed

- long.2: [121–2] -

Extant incorrect hypotheses about verb marking or verb versus adjective categories can then result in a further incorrect hypothesis about the phonological shape of the modal, which is analyzed as will be. Some learners (such as Ludwiga J.) persist with their hypothesis, while others (like Krystyna B.) modify it. In the case of a learner like Andrzej J., where will be occurs

only with copular adjectives, the incorrect hypothesis may have never been formulated.

The fact that unanalyzed will be is much more common amongst the Polish speakers may also be amenable to explanation. It has been noted in 4.7 that suppliance of the copula is much lower amongst Vietnamese speakers than amongst Polish speakers. Given this, and given that will be is a form of the copula, it is probable that Vietnamese-speaking learners pay less attention to the structures that appear to provide many Polish speakers with their first analyzable input for will.

4.9.5 'Would':

As indicated in the tables, for most informants who use it, would appears only in the formulaic verb phrase would like; no evidence exists for supposing that would in these tokens has independent lexical status.

There are, however, a number of exceptions. Canh, in the first interview, uses would in a manner that may not be formulaic, or at least represents a spreading application:

> would you tell me...?
>
> - canh.1: [814] -

> I would say he's.../...gooz man
>
> - canh.1: [2149] -

Andrzej J., in the first interview actually produces a future-in-the-past token of would:

> we decided, er...er, to...leave Poland and, er...if, er...it...would be possible...to come to Australia
>
> - aj.1: [261–2] -

In the second interview, on the other hand, all the tokens of would he produces appear with like, albeit with more variation in structure and discourse situation than for other informants. Ludwiga J., in the second interview, produces tokens of would be, rather than would like, as does Mieczyslaw M. This indicates that a gradual spread of environments is taking place.

Nevertheless, structures involving would in non-stereotyped environments are barely apparent. There are no full counterfactual conditionals in the data. Nor does would appear as a modal denoting past imperfectivity. Semantic complexity and relatively low functional load are sufficient to explain this. It might also be remembered that as a form with many functions, would constitutes a classic problem for the learner. As we will see, there may even be some evidence that would is avoided.

4.9.6 'Could':

This modal is restricted to a small group of informants, and is produced infrequently. Ewa S., in the first interview, provides an instance of could in a conditional structure, but as an echo of the same structure in the interviewer's question. Spontaneously, she uses could as the past form of can. It is this latter usage which characterizes the tokens produced by Andrzej J. (both interviews), and Mieczyslaw M. (second interview):

I could, er... er, to read in Poland

- aj.1: [343] -

Likewise for Ludwiga J., where the modal does not appear to be fully established as the tense carrier:

can... I could saw... see?... I could saw... I could sa(w)?

- lj.2: [671] -

Amongst the Vietnamese speakers only Long and Canh use could. The former produces only one token, which is somewhat confusing, but where the most probable interpretation is as a past form. Canh, on the other hand, seems quite prolific. The most frequent use of could is in conditional structures, both explicit and implied:

I cou(ld) say, ah, to you, ah... anything

- canh.1: [2325] -

I couldn('t) live in there... I am still young

- canh.1: [384] -

If you don't... don't, ah, don't test it you couldn't sell it

- canh.1: [756] -

Could also possibly appears as the past form of can:

I could(n't) remember a lot my history course

- canh.1: [1411] -

However, the uncertainty of the above example, as regards time reference, and of one of the conditionals cited above, as regards modality, makes it possible that for this informant can and could are allomorphs. (The second interview provides no evidence to the contrary). This would be a classic case of having the form before the function has been acquired. There is other evidence of a more than normally pronounced tendency for this situation to occur in the case of Canh [3].

In this connection, it is worth noting that Canh is distinguished by the size of his lexicon. He might not unfairly be described as a lexical magpie. This unusual sensitivity to words may mean that he is more likely than the majority of learners to remember (and use) function words for which his grammar does not have a place. Thus the categorial vagueness and consequent incoherence which characterize a good deal of this informant's output may be in some measure a product of the overdeveloped state of his lexical faculties.

If there is no real functional distinction between can and could in the output of Canh, then could, as it appears in the data, functions solely as a past form of can, and is not used in expressions of conditionality.

4.9.7 'Should':

Should is used by all those who produce it as a modal of obligation:

I...er...should.../...speak?

- es.1: [155–6] -

Em...my friend said I shoul(d) put some, ah

- vinh.2: [208] -

One informant, however—Phuc—uses should to express several other functions. These include probability, and implicit and explicit conditionality. The range of functions is displayed in the following examples:

they say that...we should work har(d)

- phuc.1: [441] -

to buil(d) a hall li(ke) that should take you a few mon(ths)

- phuc.1: [960–1] -

[3]Cf. Canh's morphology, where "-ing" and "the" function respectively as verbal and nominal morphemes

I should like to send them... Christmas cards

- phuc.1: [1032] -

I hope.../...I should have another zhob

- phuc.1: [72] -

If I have no responsibiliti(es) I thin(k) I should be very happy to go to school, you know?

- phuc.2: [333–4] -

Most of the tokens produced by Phuc involve acceptable uses of should. In some, however, such as the last example given above, would is probably the modal a native speaker would chose. But the only tokens of would produced by Phuc are formulaic. It is interesting to speculate why Phuc has chosen to develop should rather than would to express notions such as counterfactuality. The choice may have been arbitrary, or possibly motivated by some pedagogic conception of should as a prestige form. Or perhaps the complexities of would, with its dual role as modal and aspectual functor led to the selection of the more categorially straightforward should. Whatever the reason, it seems that the system of a learner as advanced even as Phuc can only support one such modal.

4.9.8 'Have To':

In contrast to its semantic near relative must, have to is neither as widespread nor as frequently produced.

Amongst the Vietnamese informants, the form is used by Sang, Tam, Canh, Long and Phuc. Of these, Sang, Canh and Phuc, who each produce a single token, use the form in a straightforward way, as in this example from Phuc:

I have to take care of my sister(s)

- phuc.1: [741–2] -

This is unquestionably a modal of obligation.

Tam and Long present problems, however. As already mentioned in 4.9.3 above, in the output of Tam all cases of have to are sub-cases of the locution have to learn, which appears to be an idiosyncratic form of learn for this informant. Thus, Tam can be excluded from consideration. The case of Long is similar, but not quite so clear cut. Like Tam, he uses have to learn (twelve tokens). He also produces locutions of a similar kind with related verbs:

but I have to read it I understand

- long.1: [1173] -

There are about four such examples. As well as these he produces tokens where have to appears to be an allomorph of have—once again, in about four cases:

they have to been here a long time

- long.1: [412] -

(These cases could also be adduced as further evidence of modals used as tense markers). There are also tokens where have to may be a modal of obligation, such as the following:

I have to send a letter for you

- long.1: [665–6] -

Nevertheless, such tokens are never unquestionable; the example just cited may be an attempt at a perfect, for example, like the one above it. The upshot is that it is not clear whether Long produces tokens of have to as a modal of obligation or not.

Amongst the Polish informants, tokens of have to are produced by Krystof S., Barbara B., Ewa S., and, possibly, Andrzej J. All of these are in the second group of interviews. These informants provide one token apiece, and, as the examples below would suggest, the form is not very spontaneously produced:

Yes, I know I must... I... had to learn English

- bb.2: [19] -

but, er... er, now I mus(t), uh... I have to, em, Canberra

- ks.2: [228] -

Presumably these learners have had their attention drawn to the more usual have to since the time of the first interview.

Compared with must, then, have to is not very frequent at all. This state of affairs would seem to be the inverse of that found in the output of native speakers, who seem to show a marked preference for have to or have got to. This apparent anomaly appears to provide further evidence for the form-function principle. The argument is as follows. Must is a form

with an unambiguous function, that of expressing obligation. Have, on the other hand, is a form with various functions. It is a lexical verb signifying possession, an auxiliary in perfect structures, and, in combination with the complementizer to, a modal of obligation. Learners, in accordance with the principle that a form should have a single function, select must to express obligation, despite its relatively lower frequency than have to in native speech itself. It should be stressed that must occurs with very high frequency in the output of informants who have had virtually no formal instruction (for example, Long), and cannot therefore be attributed to classroom input.

4.9.9 'May':

May is used by only three informants. Jerzy B. uses it to indicate possibility, and Phuc uses it, non-standardly, to indicate capacity.

> Yes, yeah after this school, eh, I... I m... I may... I may... er... le... learn, eh, at, eh... at university
>
> - jb.2: [820–1] -

> since I may helf my friends
>
> - phuc.1: [607] -

Ewa S. uses may in an idiosyncratic combination with will, to indicate possibility:

> She will may, c... come... to Australia yes... yes
>
> - es.2: [637] -

4.9.10 Morphology

From the point of view of morphology in the structures involving modals there are a number of observations worth making.

Firstly, there is some *-ing* marking of verbs after can and must, particularly the latter. Examples of verbs marked with *-ing* after must are to be found in **is.2**, **zj.2**, **jr.1**, **jr.2**, **bb.1**, **lj.2**, **mm.1**, **tam.2**, **long.1**, and **long.2**. Examples of verbs marked with can occur in **ks.2**, **kb.1**, **kb.2**, **canh.1**, and **phuc.1**. It is not really clear why must should figure so much more than can in this phenomenon. One possible explanation lies in the probability that learners get less corrective feedback on structures involving must from native speakers—who tend to use the more idiomatic have to in its place—than they do for structures with can—where there is no such

alternative. Another possibility has to do with the use of must as a non-standard tense marker, as discussed in 4.9.3 Verbs preceded by must and marked with *-ing* could perhaps be seen as constituting learner experiments with simultaneous lexical and morphological marking.

A further phenomenon which was noted was that must seemed to occur without any surface subject with some regularity. In the first round of interviews, Barbara B., Andrzej J., Ludwiga J., Tam, and Canh produce tokens of subjectless must. Since no counts exist for non-realization of subjects with other types of verbs, it is difficult to say whether this is merely a reflection of a more general state of affairs or constitutes a special case.

4.9.11 Summary

In summary, as Tables 4.31 to 4.34 show, there is a fairly strong implicational order for the modals—particularly for the Vietnamese speakers, as well as a reasonably clear correspondence between use of a particular range of modals and assessed oral proficiency. Must, will and can appear to constitute the first group of modals to be acquired. The second group is comprised of could, should and have to. A third group, which is hardly in evidence except in the most restricted set of verbal environments would be would and may. The similarity between the predicated orders of acquisition for German and English modals suggests that universal principles of semantic markedness may play a part in determining the learning of modals.

4.10 Negation

4.10.1 Some Methodological Problems

Tables 4.35 to 4.40 provide the figures for an outline of the system of negation as it appears to develop over the cross-sectional sample. The tables are divided into seven or eight basic categories.

Within these categories, separate counts for no, not, and any other relevant negator, such as don't, are maintained. It has to be said, however, that in the process of transcription it is sometimes difficult to distinguish between tokens of no and tokens of not. There are different reasons for this. The negator may occur before a word beginning with */t/* or */d/* (for instance, as in no(t) too much), which makes it difficult to decide whether it terminates with a consonant or a vowel. In the case of the Vietnamese speakers, the negator may be terminated with a glottal stop, rather than an alveolar stop. Glottal stops are a legitimate allophone of final */t/* under certain conditions in English, but of course it is difficult to know whether a glottal stop produced by a Vietnamese speaker, although it results in a word that sounds more like not than no to native speakers, is merely fortuitous or is actually evidence that the learner has acquired a viable and system-

Informant	is	ks	zj	jb	jr	ka	bb	es	aj	kb	lj	mm
must		5	3		5	5	18	8	2	8	18	12
will				2*	1	5		2	8	6	5*	8
can						1		7	3	11	12	4
can't							1*				6	
would					1F		10F		5*			
I'd					1F			5F				
could								3	1			
couldn't									1?			
should								1				

Table 4.31: Distribution of Modals: Polish Informants—Interview 1

NOTES:

1. All counts suffixed with an "F" indicate that the tokens involved are formulaic. This applies here to "would", which always appears in the phrase "would like".
2. In the case of **jb** both tokens of "will" occur in the chunk "will be".
3. In the case of **bb** the token of "can't" is echoed.
4. In the case of **aj** four of the five tokens occur with "like".
5. In the case of **lj** "will" occurs in the chunk "will be".

atic distinction between the two morphological forms of the negator. These remarks should be borne in mind when examining the counts. In any case, the acquisition of a distinction between no and not involves a morphological, rather than a syntactic rule, and is really a separate matter. Having said all this, there is nevertheless, as we will see, evidence in the data on negation that just such morphological distinctions do tend to develop along with the syntactic ones.

4.10.2 Categories of Negation

The categories themselves are the following:

Standalone or Holophrastic 'No' and 'Not'

This is the negator used by itself in answer to a question, or to negate some proposition originating from the interlocutor, or, occasionally, in correction of a previous proposition by the speaker himself.

> GB Hm, mm...and you never studied English at school? ...in Poland?
>
> I No...never

Informant	is	ks	zj	jr	ka	jb	bb	es	aj	kb	lj	mm
must	6	8	2	7	2	5	22	9	3	10	18	24
will			1	5	2	1	20*	10	14*	6	37*	5
won't									1	1	2	
we'll					1			3				
what'll								1				
can	1	3				1	3	3	14	14	15	2
can't						1		4		12	5	
I'd				1F				1F	1F			
would									9F		2	1
could									2		1	
couldn't												1
may						2		2*				
should								4				
shouldn't								4				
have_to		1?					1	1	1?			

Table 4.32: Distribution of Modals: Polish Informants—Interview 2

NOTES:

1. All counts suffixed with an "F" indicate that the tokens involved are formulaic. This applies here to "I'd", which always appears in the phrase "I'd like".
2. In the case of **bb** all twenty tokens of "will" occur in the chunk "will be"; in many cases the resulting utterances are non-standard.
3. In the case of **jr** four tokens of "will" occur in the chunk "will be"; in one case the resulting utterance is non-standard.
4. In the case of **es** the tokens of "may" both occur in the phrase "will may", which is this informant's semantic equivalent of "might".
5. In the case of **aj** eight of the fourteen tokens of "will" occur in the chunk "will be"; in all cases the resulting utterance is standard.
6. In the case of **lj** "will" always occurs in the chunk "will be". Likewise, both tokens of "won't" involve "won't be".

Informant	Van	My	Duc	Dung	Minh	Sang	Hoa	Vinh	Tam	Canh	Long	Phuc
can			5	1		6	5	18	6	7	26	17
can't				2		3	1	7	12	7	16	7
must					1	3	1	6	9	7	40	5
will					1	1		9	4	7	3*	23
would		1F				5F	2F	1F		4	3F	2F
have_to									13*	1	27*	2
could										7	1	
couldn't										8		
should										2		15*
may												1*

Table 4.33: Distribution of Modals: Vietnamese Informants—Interview 1

NOTES:

1. All counts suffixed with an "F" indicate that the tokens involved are formulaic. This applies here to "would", which always appears in the phrase "would like".
2. The asterisked counts of "have to" all involve substantial numbers of cases where the token may not be a modal of obligation, or may be functioning as a reduplicated companion to "must" to indicate very strong obligation.
3. In the case of **long** all tokens of "will" occur in the chunk "will be".
4. In the case of **phuc** "should" appears to somewhat of a catchall modal for this informant. It is used to express obligation, probability and in counter-factuals. This probably accounts for the appearance of "would" as a formula only. "May" is used to indicate capacity, rather than probability or permission.

Informant	Van	Duc	Sang	Minh	My	Hoa	Dung	Vinh	Tam	Canh	Long	Phuc
can		4	2			2	1*	1	27	14	4	12
can't			10	4				1	16	3		11
cannot										1		
must			1		1	1		9	19*	3	7	1
will			1		1		1*	7	3	1	3	7
won't							1*					
have_to			1						1	1	2	1*
would							1F			1?		2F
should								2			2*	20*
could										1		
couldn't										3		

Table 4.34: Distribution of Modals: Vietnamese Informants—Interview 2

NOTES:

1. All counts suffixed with an "F" indicate that the tokens involved are formulaic. This applies here to "would", which always appears in the phrase "would like".
2. In the case of **duc** the asterisked tokens all occur in the same utterance, while the informant is searching for the right form, which is "won't" in this case.
3. The asterisked counts of "have to" all involve substantial numbers of cases where the token may not be a modal of obligation, or may be functioning as a reduplicated companion to "must" to indicate very strong obligation.
4. In the case of **long** "should" is used to express probability.
5. In the case of **phuc** "should" appears to somewhat of a catchall modal for this informant. It is used to express obligation, probability and in counter-factuals. This probably accounts for the appearance of "would" as a formula only. The token of "have to" is actually "have got to".

- jb.1: [29–32] -

... er, eleven years... eleven, no, no... (i)s two, thr... eight!... eight year(s) in, t... another factory

- jr.1: [121–2] -

At times, standalone no may have a holophrastic function, in that it condenses what would normally be a more extended negative proposition into a single word. This, obviously, is more likely to occur with speakers whose systems of negation are not very developed.

GB Mm... yes... ah, were you in Poland when Solidarity was formed?

I No.

GB You were in Austria?

I Er... no, no... er... in, er... ch... ch... I was in Poland S, Soli, Solidarity?

- jb.1: [345–53] -

It is interesting in this regard that the frequencies for standalone no tend to peak in the middle of the *ASLPR* range and to decline thereafter, despite the fact that informants higher in the range have significantly higher total word counts and could therefore be expected to produce more tokens of no as a side effect of this. Table 4.40 gives the figures for no expressed as a percentage of the total number of words used—cultural factors might have something to do with the slightly lower figures for the Vietnamese speakers. It is a little difficult to know exactly how to interpret these declining frequencies. One possibility is that standalone no does occur in a holophrastic capacity to a significant extent beyond a certain level of development. Yet another is that standalone no is not used holophrastically to such a significant degree, but that misunderstandings and communication breakdowns require a good deal of correction, reformulation, and therefore frequent negation on the part of the speaker. It would be interesting to investigate these proposals further.

Phrasal Negator 'No' or 'Not'

This category includes all those cases where the negator occurs in conjunction with an adjective, adverb, quantifier, noun phrase, prepositional phrase, etc., where these phrases occur in isolation, or from intonation and/or context are evidently not part of a verb phrase. Examples would be:

... in hostel, er, no... no longk

- zj.1: [786] -

GB Ah ha... did he speak English very well?
I Na... no very well

- zj.1: [489–91] -

Drink beer, yeah, drink beer, too much, now no, now not much... not too much

- dung.1: [127–8] -

boat not ship

- dung.1: [166] -

not in the city

- tam.1: [499] -

Such tokens can be quite acceptable products of ellipsis, of course, and are frequent in native speech. On occasion, however, they are the result of failure to insert the copula, or failure to insert both the copula and a subject. For reasons which will become clear, tokens of negated phrases within a verb phrase—that is, preceded by a verb (generally the copula)—are not counted as examples of phrasal negation.

Negative Quantifier 'No' or 'Not'

This category includes those cases where the negator functions as a quantifier. (Quantifiers in general are dealt with in more detail in 4.14.5). Also grouped with negative quantifiers is the small group of idiomatic (and formulaic) phrases which includes no good and no more, since in so far as these phrases can be parsed the no component is in fact a quantifier. Initially, these were to be eliminated from the counts for negative quantifiers, and listed separately. However, a check revealed that informants who produced such idioms also produced non-formulaic tokens of no as a negative quantifier, and that separate listings would have not altered the overall counts very greatly. Examples of no as a negative quantifier are:

my factory no job(s)

- dung.1: [589] -

No good... every day goulash

- zj.1: [770] -

Preverbal Negators

Preverbal negators include no, not and various forms of *-n't*, such as don't, can't and haven't. Some examples are:

No understand

- ka.1: [490] -

...we drink, em...vodka and, er, we, em, not eat

- mm.1: [470] -

in Vietnam I, er...no study English

- minh.1: [20] -

I don('t) have any re(la)tive(s) in Australia

- sang.1: [24] -

I didn't, er, travel before my...er, leave Poland

- aj.1: [580] -

I can't work

- bb.1: [157] -

Postverbal Negators

Postverbal negators include no, not, and some forms of *-n't*. There are several problems in classifying postverbal negators.

With the *-n't* morpheme, there is a problem in deciding whether the morpheme is really analyzable as a contracted form of not, in which case it would classify as a postverbal negator, or whether it is part of a word with no internal structure, as it obviously is in early cases of don't. The best approach seemed to be very conservative in this matter and to only count tokens of *-n't* as evidence of post verbal negation if there was evidence of actual decomposition of the word into element plus negator elsewhere in a particular informant's output, in the case of those words which themselves functioned as preverbal negators (that is, don't, can't, etc.). In the case of words which were not used as preverbal negators (for instance, isn't or wasn't), these were tentatively classed as tokens of postverbal negation, provided they did not occur in formulas such as question tags. The result was

that there are very few clear tokens of postverbal negation represented by words ending in *-n't*. As outlined above, such a decision may seem both arbitrary and unnecessarily restrictive; more detailed examination of the data, however, will show that it is probably well motivated. With the no and not morphemes, there are problems of an almost opposite kind. In the majority of cases of postverbal negation the verb concerned is the copula. Here the difficulty is in deciding whether the negator should really be grouped with the verb and assigned the status of a verbal negator, or whether it would more properly be considered as a phrasal negator. Examples of possible postverbal negation would be the following:

Europe ist... er... not safe

- es.1: [141] -

we have not enough materials

- phuc.1: [941] -

The difficulty with the analysis is to know whether to treat a structure such as this as a case of *is not* || X or *is* || X. This problem is discussed in more detail further on. In general, the approach has been to treat the negator as verbal, unless evidence from pauses or intonation suggests otherwise. More evident examples of postverbal negation than the ones cited above are:

is not problem for me

- bb.1: [574] -

she was not happy

- phuc.1: [333] -

I... will no, er... work

- ka.1: [439] -

Anaphoric 'No' or 'Not'

Anaphoric negators refer back to some proposition advanced at an earlier point in the discourse; they are relatively uncommon in the present data. Examples would be:

I'm not sure... it is true, maybe not

- lj.1: [386] -

you can get job like me now or not?

- long.1: [785–6] -

she like but she can('t)

- dung.1: [872] -

Question Tag 'No'

This is not an important category, but it seemed worth including. An example would be:

You married, no?

- ks.1: [970] -

Verbal Complement 'No'

These could possibly be treated as a subset of anaphoric negators. The examples were tabulated mainly because they provide evidence of how morphological distinctions that are well established in one area will not necessarily be made elsewhere. An example of verbal complement no is:

No, I think no

- lj.1: [443] -

4.10.3 A Postulated Developmental Sequence

Although it is necessary to exercise some caution when postulating a developmental sequence on the basis of cross-sectional data, Tables 4.35 and 4.40 suggest a sequence which is very much in line with that described in other research [27]. For the moment, the morphological distinction between no and not will be disregarded. Some points about this distinction will be made when the basic sequence has been described.

Holophrastic Negation

The first step in the acquisition of a system of negation comes with the use of standalone or holophrastic no. For early learners the semantics of no, as for many other words as well, are somewhat wider than in the target. Thus no serves not only to contradict propositions made in the discourse, but also to indicate incapacity—for example, incapacity to answer questions.

As has already been pointed out, the relative frequency of standalone no declines as oral proficiency increases and the system of negation develops. Two possible explanations for this are offered in 4.10.2 above.

At this point, the locution don't know appears. At this stage there can be very little doubt that this is not monomorphemic. As with no itself, don't know, has a wider semantic field than in Standard English, serving to indicate incapacity as well as lack of knowledge.

MJ .../...do you have a job?

I [LONG PAUSE] I...I don't know

- van.1: [30–32] -

The case of don't know is a very clear example of the acquisition of the form before the function—a phenomenon which, as we will see, is in evidence elsewhere in the acquisition of negation, and is a productive aspect of the acquisitional process.

Phrasal Negation with 'No' or 'Not'

The next stage of negation, as evidenced from the distribution patterns in Tables 4.35 and 4.37, is the preposing of a negator, either no or not, before some other element. (Don't cannot yet be classed as negator, since it still only occurs in monomorphemic don't know and don't understand). At this particular point of a learner's development, the categorial status of the negated element is often rather vague. English does not have an abundance of morphological markers to indicate categorial status, and in the elliptic utterances which are produced by learners at this stage of their development there are insufficient elements for parsing to provide unambiguous categorial labels. Examples of this categorial vagueness would be:

GB .../...did you work, at all, or...?

I In Aus(t)ria?

GB In Austria.

I No, no...no work

- jb.1: [155–161] -

MJ ...were you in the army?

I I no army

- my.1: [154–6] -

In any case, the categorial identity of the negated elements is only crucial to the analysis if some categories of words appear to be negated before others. There is some evidence from child longitudinal studies that verbs are negated by neg-preposing later than, say, noun phrases or adjectives [185]. This conclusion is neither contradicted or strongly supported by the present study. One informant, Irena S., produces tokens of phrasal negation, but not verbal negation. Other informants produce both, with phrasal negation tending to be more frequent than preverbal negation with the informants lower on the *ASLPR* scale. If there is indeed a time differential for these two types of negation for adult learners, it is probably slight.

Categorial vagueness itself, as a major phenomenon, does not persist for very long, as learners begin to incorporate words that are recognizably members of the main grammatical categories into their vocabularies. Examples of phrasal negation, involving adjectives, adverbs, quantifiers, noun phrases and prepositional phrases are given above in 4.10.2 Noun phrases and adverbs are the most frequently negated categories for the early learners.

At the stage of phrasal negation, or perhaps slightly later, no also appears as a negative quantifier. This is not surprising, since the structural description of negative quantification with no and phrasal negation is the same, although the semantics are different. Some of the tokens of no as a negative quantifier are represented by phrases like no good and no more, which are, of course, formulaic.

Preverbal negation is also in evidence at this point, constituting either a subset of cases of phrasal negation or, just possibly, a later extension of the neg-preposing rule to the category of verb. The first definite preverbal negators are no and not. Examples are given in 4.10.2 above.

Preverbal Negation with "-n't" Forms

While the first appearances of don't are as part of formulas, by about 1– on the *ASLPR*, there is some evidence that this morpheme is beginning to establish itself as an alternative preverbal negator. Table 4.39 shows that Duc uses don't in several verbal environments. Amongst the Polish informants, Jerzy B., while producing the majority of his tokens of don't with know and understand, produces one token with have. This sort of evidence shows how a new morpheme gradually assumes a definite lexical identity and function, and provides a very important insight into the learning process. Tables 4.36, refneg:453 and 4.38 provide details of the various environments for don't for all the informants in the study. While know and understand remain the dominant environments for most speakers, it can be seen that there is general increase in the range of verbal environments as one proceeds up the *ASLPR*. Regarding the final step of the acquisition of the auxiliary do as the carrier for the negator, however, there is no definite evidence in the present data, from negation itself at any rate, that don't is further

decomposed into do and not. The implications of this point for teaching will be discussed shortly.

A variant of don't, doesn't, also begins to appear in the speech of the Vietnamese informants, though not the Polish speakers, by about level 1. The majority of instances for all speakers involve the phrase doesn't matter. Where other verbs occur with doesn't, the form does not necessarily serve to mark the third person singular, as it does in Standard English (in Long's case it always occurs after they) and if it does mark this person, does not do so consistently.

There are other preverbal negators besides don't and doesn't. These are detailed in Tables 4.36, 4.37 and 4.38. Once again, these increase in number, and to some extent frequency of occurrence, across the **ASLPR**.

These preverbal negators with an -n't form present somewhat of a paradox. In native speech, such forms would not be classified as preverbal negators at all: they would in fact be examples of postverbal negation, with their forms resulting from the contraction of the postverbal negator not and its subsequent attraction to the auxiliary or modal.

As we have seen, there is overwhelming evidence that don't cannot be analyzed in this way when it first enters learner speech. Far from being analyzable into two components, it is not even a discrete element itself when it appears in the formulaic don't know. When don't becomes a discrete element, non-standard patterns of usage, such as in the example below, indicate that it is monomorphemic, and still cannot be decomposed into do and not:

> I want to make, er, to photograph again... in Australia... but, er, but don't licen(ce)
>
> - minh.1: [412–3] -

> Doesn't freedom... freedom... not freedom
>
> - dung.1: [154] -

For these reasons, it is safe to ascribe to don't the role of a preverbal negator, and not to treat its use as evidence of postverbal negation until such time as informants produce clear evidence that in their grammars it is a decomposable form—say, by producing tokens of do not, or perhaps by their use of do in questions.

The question that now has to be asked is whether the other -n't forms also have the same monomorphemic status as don't when they first begin to appear in learner speech.

Turning to Tables 4.36, 4.37 and 4.38, the first point is that in the output of the Polish speakers haven't, wasn't, isn't, and the dubious token of weren't,

while they can occur as preverbal negators as auxiliaries, do not do so in the data. They have been left in the table because of their possible significance as evidence for postverbal negation, of which more later.

The next point concerns haven't in the output of the Vietnamese speakers, and illustrates some of the problems referred to earlier in determining whether or not certain words have an internal structure. As can be seen from Table 4.38 a number of Vietnamese speakers use haven't. One of these, Vinh, uses it as a main verb, rather than as a preverbal negator, and can be excluded from the present discussion. The principal usage for the other speakers is in the locution haven't got. This covers all the tokens for Tam, Canh and Long, and three out of four tokens for Dung, leaving Sang the odd man out. Further investigation reveals that it is only Sang who uses have as an auxiliary as well as haven't, which serves to differentiate his case still further. The other informants, in declarative statements where have plus past participle would be standard—and these all involve got too—merely produce the participial form as a main verb. (This point is discussed further in 4.2). The only form of auxiliary have these informants produce, then, is haven't, and then almost always in the locution haven't got. The following examples are typical of the pattern of occurrence of got:

he got, er... er, little boat

- long.1: [174] -

we Army before we haven't got any money

- long.1: [187] -

There is an obvious parallel in this pattern of use of haven't in haven't got with that of don't in don't know.

The pattern is not exactly equivalent, in that informants when they first begin producing don't know, do not have do as a discrete element, whereas the learners who use haven't in the way described above already use have. Nevertheless, there must obviously be considerable doubt that haven't used in this way is anything but a specialized form of preverbal negator.

This doubt is strengthened when we look more closely at the patterns of use for have and haven't. Ostensibly, these words constitute an *affirmative/negative* pair. This in itself is not a guarantee that a speaker will isolate the *-n't* morpheme as a negator in his or her grammar, but it certainly makes such a move possible. The fact is, however, that two elements in this pair do not have the same categorial status. That is, the affirmative element have is used only as a main verb (except by Sang, who has already been noted as an exception), while the negative element haven't is used exclusively before main verbs, as either negator or auxiliary. The pair, then, is not really a

pair. When it functions as a main verb have is negated by don't, not the *-n't* morpheme, or uncontracted not. And as an auxiliary (in the cases under discussion) have simply does not appear at all. This makes it almost certain that for the speakers in question, haven't is monomorphemic, and therefore should be classified as a preverbal negator exactly as for early don't.

The case of haven't in the output of Dung, Tam, Canh and Long indicates how careful it is necessary to be in ascribing target structures to target-like forms. It also illustrates how a form which could ostensibly be classified primarily as an example of postverbal negation which just happens to be functioning as a preverbal negator in its matrix structure is in fact a preverbal negator pure and simple. It is possible that other forms, such as didn't, which in the target constitute examples of postverbal negation actually first enter the learner's system as formulaic preverbal negators.

Amongst the other forms of *-n't*, can't is probably the most likely candidate for a form with genuine internal structure. Nearly all of the informants use both can and can't, and, as mentioned above, the existence of such a pair should facilitate the isolation of the *-n't* morpheme as a negator, although the mere existence of the pair is not evidence that this has actually happened. Even here, however, especially in the case of the Vietnamese speakers, the picture is clouded by the non-production of the crucial final */t/*, and by the fact that the can/can't distinction is marked in the vowel in the target and by some learners, and not only by the ending.

For these reasons, most of the tokens of *-n't* forms in Tables 4.36, 4.37 and 4.38 have been treated as examples of preverbal negation, unless there is strong evidence to the contrary.

Preverbal Negation—The Spread of "-n't" Forms

To return to the description of the sequence itself. Once the *-n't* forms begin to emerge from their formulaic contexts, as detailed in Tables 4.36, 4.37 and 4.39, there is a general decline in the use of the non-standard preverbal negators no and not. This appears to take place at about level 1 on the *ASLPR* for both groups. For the Vietnamese speakers, the regularities are striking: of the last five informants only Long produces any tokens of preverbal no or not. In the case of the Polish speakers there is some persistence of not as a preverbal negator, though the percentages are low.

Postverbal Negation

The next probable stage after the inception of preverbal negation is postverbal negation. Some of the difficulties in classifying tokens as definite cases of postverbal negation have been discussed in 4.1.4 above.

For the Polish speakers postverbal negation appears somewhat earlier than for the Vietnamese speakers, and is more frequent at all stages. To

find an explanation for this we have to remember that standard, and (in this data) most non-standard, postverbal negation is restricted to a particular class of verbs. These are be, have, modals and auxiliaries. The most frequent of these verbs is the copula. As is detailed in 4.7

Vietnamese speakers have a far lower rate of copula insertion than Polish speakers, and this would obviously affect their opportunities for postverbal negation. Structures which would normally provide occasion for postverbal negation would, minus the copula, turn into examples of phrasal negation. These can in fact be found scattered throughout the data from the Vietnamese speakers.

it not much bad thing

- vinh.1: [626] -

I no happy here

- long.1: [303] -

There is another curious difference here, and that is that while Polish informants who use have as a main verb almost always produce it in negative form as haven't, the Vietnamese informants, with the exception of Vinh, choose to negate it with don't. One explanation for this could be the Vietnamese speakers' lack of a well-founded paradigm for postverbal negation with no or not—they may simply stick to preverbal negation as much as possible. Another possible explanation lies in the differential patterns of exposure to English for the two groups, with the Vietnamese speakers, having on the whole received more naturalistic exposure, using the form most frequent in native speech. If this latter explanation is valid, then it would be a case of type of input having a significant effect on the development of syntactic rules [185].

The Case of "Isn't"

In the context of postverbal negation, there is one more interesting phenomenon to be observed. This concerns the absence of a form that one would have expected to be quite common—namely isn't. Only one informant uses isn't in the first round of interviews, and more than once in the second. This is Ludwiga J. (the others are Ewa S., Mieczyslaw M., and Tam). The predicates are quite varied:

it isn't only this branch

- lj.1: [273] -

it isn't shop... what is this?

- lj.1: [329] -

It is not clear why Ludwiga J. is such a prolific user of this particular item. On the other hand, a possible explanation for its non-use by other learners can be put forward. This is as follows.

Postverbal negation of the copula follows naturally from the extension of negated phrases by the addition of a subject and copula, or the copula alone. Thus a learner who already produces structures like not every day or my house, er, no(t) in city will produce examples of postverbal negation almost as a side effect of the developmental step of using the copula. Having arrived, as it were, almost fortuitously at a satisfactory means of negating the copula, the learner has no need to search for a specifically negative form of is, was, or any other form of the copula. This is in contrast to other members (with the partial exception of have) of the class of verbs that behave like be in regard to placement of the negator, whose semantics and syntax require active decisions on the part of the learner as to how they will be negated—consider what happens if can is tacked on to an already negative verbal complement.

The case of isn't may in fact be another illustration of the form-function principle—the form is so conspicuous by its absence that one almost suspects a constraint against it, on the grounds of its redundancy.

The case of isn't—as well as the other *-n't* forms—has rather drastic implications for any teaching practice based on the assumption that contracted forms can be taught by demonstrating their derivation from uncontracted ones. From the evidence presented above, in most cases it is the contracted forms which are learnt first, and subsequently decomposed. In the event of an uncontracted structure being acquired before the contracted form, as with is not and isn't, the contracted form will probably not be learnt at all. Right throughout the data, there are very few duplications of this kind. For the numerous tokens of can't, there is only a single token of cannot (**canh.2**). For other *-n't* forms the situation is hardly better: there are six tokens of have not (four in **bb.1**, and one each in **bb.2**, **kb.2** and **phuc.1**), and two dubious tokens of do/does not (**duc.2** and **vinh.1**).

The Sequence for Negation—Summary

This completes the description of the major steps in the development of a system of negation. In the development of a target-like system, of course other rules would have to be learnt, and the rules already described would have to be consolidated. One area of importance not dealt with here concerns the rules governing the behaviour of quantifiers in the scope of negation—that is, the suppletion of words like some into any. These matters are touched upon in 4.14

The basic finding in relation to the rule of suppletion is that it is language dependent (and therefore possibly not developmental), with the Vietnamese speakers performing better in this area than the Poles, and beginning to exercise the rule after level 1 on the **ASLPR**. Regarding the features tabulated in Tables 4.35 and 4.37 and not yet discussed, namely Anaphoric Negation and Verbal Complement, little needs to be said. From the small amount of evidence available these features emerge largely after *ASLPR* levels 1 and 1+ respectively, but could be subject to individual variation.

4.10.4 Morphology

Having completed the major part of the survey on negation, we are now in a position to make some comments on the development of morphological distinctions within the system—namely, the distinction between no and not.

In the introductory part of this section, some of the problems relating to the transcription of these two items were discussed, and the limitations pointed out should be kept in mind during the subsequent discussion.

The first point to make about the negators no and not is that the phonetically simpler form no must be considered to be the prototypical form. There is evidence to support this phonological prediction. For the two usages where no is the actual target form, standalone no and negative quantifier no, there are very few tokens of not, whereas for the remaining usages, where not is the target form, there are numerous tokens of no. The hierarchy of difficulty for the two forms is therefore quite clear, and does not work in favour of the learner with the exception of the two usages cited above.

Given this, it is not surprising that no is initially more frequent than not in preverbal negation (where both forms are non-standard), postverbal negation (where no is non-standard), and phrasal negation (where no is again non-standard). Tables 4.35 and 4.36, however, show a reversal of this trend as one moves across the *ASLPR*.

Thus, in phrasal negation, for both language groups, no declines and finally disappears as not increases.

In postverbal negation amongst the Polish speakers, where the distribution of tokens is wider, no declines even more rapidly than in phrasal negation. Since the rule for postverbal negation may be easier to acquire than that for phrasal negation, this suggests that there is some interaction between syntax and morphology here. In postverbal negation for the Vietnamese speakers, the situation is not so clear because the phenomenon is more restricted in distribution, but by the stage postverbal negation appears not is only morpheme for one informant, and much the more frequent for the other.

In preverbal negation for the Vietnamese speakers, both no and not cease abruptly after Sang, with the exception of some tokens in the output of Long, whose morphological profile exhibits other inconsistencies as well. With the

Polish speakers, the basic pattern of no grading into not occurs once again. This is an interesting phenomenon, given that preverbal not is non-standard anyhow. The informants involved, Krystyna B., Ludwiga J., and Mieczyslaw M., have either abandoned no entirely (except for legitimate contexts) or almost entirely, indicating that they have formed a general hypothesis about which of the two forms is likely to be correct, even in contexts such as preverbal negation, where their rule-systems are still unstable. For Mieczyslaw M. this overgeneralization even results in a case of standalone not.

As far as the morphology of negation is concerned, then, it appears that there is a definite trend towards the acquisition of not over no, where this is appropriate, in parallel with syntactic developments. As to whether there is some interdependence between the morphology and the syntax, this is somewhat of an open question. It is possible, though by no means certain, that the process of acquiring not is slower in areas where the syntactic rules are more difficult to formulate—as in the case of phrasal versus postverbal negation in the Polish data. On the other hand, there is also evidence (though not from the same informants) that developmental processes in morphology can operate independently of syntactic ones and result in "standardizing" trends within non-standard structures—as in preverbal negation in the Polish data. Results like these suggest that we can probably expect a good deal of individual variation in this area.

Given the lack of any definite links between progress in the morphology and the syntax of negation, and the likelihood of individual variation, it would probably be best for teachers not to dwell on morphology if their aim is to inculcate syntax, particularly with learners below level 1.

4.10.5 Conclusions

The broad developmental sequence in negation begins with standalone no, and formulaic don't know (at about 0 on the *ASLPR*). There follows a stage in which phrasal and preverbal negation emerges, along with further expressions containing don't (at about *ASLPR* level 0+). Following this are the first examples of postverbal negation, involving the copula in the first instance (at about *ASLPR* level 1−). Preverbal negation is characterized further at this point, or soon after, by the emergence of don't as a separate morpheme, and the use of other *-n't* forms as preverbal negators. Some of these latter forms may also constitute examples of postverbal negation. Examples of suppleted any may begin to appear soon after (*ASLPR* level 1 onwards). The decomposition of *-n't* forms into their separate elements is still not evident, however. In the area of morphology, not begins to supplant no in the appropriate contexts, with some degree of individual variation.

There are various important implications to be drawn from the acquisition of the rules for negation.

Negation is an area of learner grammar which provides a rich fund of

examples of learning as a process of decomposition and discrimination. The way in which don't enters the system as part of a larger chunk, is isolated as a negative morpheme as it spreads to other environments, and may finally be decomposed into the separate constituents of auxiliary and negator is an excellent example of this process. The way in which don't and other *-n't* forms, such as isn't, are treated in learner grammars is also instructive of how dramatically such grammars can differ from those of native speakers in respect of the same elements, and highlights the dangers of using intuitive "simplifications" as a basis for teaching. Negation also provides pertinent examples of the interdependence of the various rule-systems that make up a total grammar: the dependence of postverbal negation on copula insertion and the interrelations between interrogation and negation through shared elements such as do and any are two obvious examples.

4.11 Questions

4.11.1 Limitations

The following analysis of questions is subject to certain limitations, as regards the data collected. This is because it is difficult to obtain all the data on questions that one might want without either compiling an enormous corpus or resorting to specific elicitation procedures. In fact, a substantial amount of data on interrogation and various other less accessible facets of learner speech actually was collected during the second round of interviews (generally at the end of Interview Two) with the help of written prompts and translations, but given the informal nature of the rest of the two interviews, a decision was taken not to mix this data with the more spontaneous conversational material which constitutes the main body of the interviews, and which has provided the basis for all of the analysis so far conducted. This elicited data still awaits analysis, and should produce some very instructive comparisons with the free speech material. As a consequence of the decision to concentrate the analysis on the informants' spontaneous production, the present analysis of question formation deals only with material obtained in the informal part of the interviews.

4.11.2 Collection of Question Data

Some of this material is totally spontaneous in origin—a by-product like any other of the conversational transactions that took place. The rest is the result of a request made at the end of the first interview that the informants ask the interviewers some questions about themselves. Only the most general suggestions as to what these questions might be were provided, and the interviewees were not pressured to proceed if they appeared reluctant to do so.

Informant	is	zj	ks	jb	jr	ka	bb	es	aj	kb	lj	mm
STANDALONE												
no	16	53	27	25	41	45	12	32	2	13	11	17
not			1									1
PHRASAL NEG												
no	1	10	5	3	5	2	4	3	3	1		
not	1	3	1	1		2	4	5	6	11	20	13
NEG QUANT												
no	1	3	3		3	1			1	3	3	2
PREVERBAL												
no		1		2	9	2						
not			1		1		6			3	4	4
-n't		6F	8F	12		1F	15F	24	15	21	41	15
POSTVERBAL												
no		1?	1?			3						
not			2		2	2	24	1	6	12	9	12
-n't				1?		3		2		9	30	7
ANAPHORIC												
not							2				1	2
-n't						1		1			2	
QUEST TAG												
no			1				1					
VERB COMP												
no									1		1	

Table 4.35: Forms of Negation: Polish Informants—Interview 1

NOTES:

1. Counts suffixed with an "F" indicate probable formulaic usages of the token(s) in question. For the purposes of these tables, formulaic usages are defined as those which involve the appearance of the morpheme in question in no more than two verbal environment. (See the text for a fuller discussion of stereotyping).

Informant	van	my	duc	dung	minh	hoa	sang	vinh	tam	canh	long	phuc
STANDALONE												
no	1	9	13	19	12	26	12	7	5	22	30	4
PHRASAL NEG												
no		3?		2	1	1	2				3	
not		1	7	2	15	2	1	9	9	11	33	8
NEG QUANT												
no			5	3	1	6	5	1	2	2	9	1
not			1	2			1?					
PREVERBAL												
no		1		1	2	5	1				17	
not		3	1	1		5*	1				5*	
-n't	2F		7	14	3F	30	7	17	33	62	105	47
POSTVERBAL												
no								2			1?	
not								9		2?	1?	10
-n't							1	2				
ANAPHORIC												
not								1			1	1
-n't				1		1	1		1			1
VERBAL COMP												
no									1			

Table 4.36: Forms of Negation: Vietnamese Informants—Interview 1

NOTES:

1. Counts suffixed with an "F" indicate probable formulaic usages of the token(s) in question. For the purposes of these tables, formulaic usages are defined as those which involve the appearance of the morpheme in question in no more than two verbal environment. (See the text for a fuller discussion of stereotyping).

Informant	is	zj	ks	jb	jr	ka	bb	es	aj	kb	lj	mm
know		3	4	6			12	18	6	14	18	15
understand		3	4	4		1	2	3	4	2		
remember								1	1			
have				1								
like									2			
want											3	
think											1	
OTHER				X				X	X	X	X	

Table 4.37: Environments for "don't": Polish Informants—Interview 1

Informant	van	my	duc	dung	minh	hoa	sang	vinh	tam	canh	long	phuc
don't	2		7	3	3	28	3	8	20	27	78	25
can't				4		1	3	5	8	8	17	7
doesn't				2		1		1	1	8	3	5
haven't				4			1	2	4	11	6	
didn't				1				2			1	10
couldn't										8		

Table 4.38: Distribution of "n't": Vietnamese Informants—Interview 1

Informant	van	my	duc	dung	minh	hoa	sang	vinh	tam	canh	long	phuc
like			1	3		3		4	5	3	6	3
know			1			16	1		4	13	20	7
have			1			2	1	3			3	
understand			3		1				5		34	
want						1				3		
think										3		1
remember										3	1	
care											9	
OTHER			1		1	3		1		5	9	10

Table 4.39: Environments for "don't": Vietnamese Informants—Interview 1

Int	%
aj.1	0.1
phuc.1	0.1
lj.1	0.2
tam.1	0.2
canh.1	0.3
long.1	0.3
vinh.1	0.3
bb.1	0.5
kb.1	0.5
mm.1	0.5
duc.1	0.8
sang.1	0.8
es.1	1.0
jb.1	1.0
jr.1	1.0
ks.1	1.0
van.1	1.0
dung.1	1.5
hoa.1	1.5
minh.1	1.5
is.1	2.0
ka.1	2.0
my.1	2.5
zj.1	4.0

Table 4.40: Standalone "No" as Percentage of Total Words

Naturally some of the learners were less inhibited than others, and the amounts of data vary somewhat. A further consequence of collecting material in this fashion is that it put the informants under no real pressure—they were not there to find out things they really needed to know, and therefore they were not obliged to stretch their linguistic resources or use what skills they had very creatively. Instead, they tended to stick to the predictable, which probably accounts for the large number of formulaic questions produced.

Furthermore, the presence of large numbers of formulaic questions constitutes a problem in itself, since in an analysis of the present kind there is insufficient data for implicational scaling and it is consequently necessary to make partly arbitrary decisions about where formulas cease and rule application begins.

Nevertheless, the results obtained suggest definite patterns, and constitute a useful set of working hypotheses for further investigations in this area. In particular, it would be most interesting to test their validity against the more extensive sets of elicited questions gathered at the end of the cross-sectional study, especially as these latter questions were modelled to some extent on the original spontaneous ones.

4.11.3 Structures Analyzed

The present analysis is principally concerned with four structures. These are WH-questions and Yes/No questions, involving either inversion or *do*-insertion. Well-formed examples of the two types of WH-questions are:

What is your hobby?

- jr.1: [1302] -

and where is the winter?

- bb.1: [494] -

how long can we... er... live here in hostel?

- ka.1: [1133] -

and:

Whas do you do now?

- sang.1: [512] -

Where do you live now?

- dung.1: [734] -

When did you get married?

- dung.1: [708] -

And well formed examples of the corresponding Yes/No set are:

Is thi(s) your car?

- minh.1: [273] -

Are you a teacher?

- es.1: [1184] -

are you married?

- jb.1: [984] -

have you ever children?

- ka.1: [1200–1] -

can you he(l)p me?

- hoa.1: [1026] -

and:

Do you understand me?

- kb.1: [257] -

did you know, um...when the Vietnamese, um, firs(t) came to Australia?

- sang.1: [575–6] -

These four structures (at times referred to for convenience as *WH/INV*, *WH/DO*, *YN/INV* and *YN/DO*) should present the learner with learning tasks of differing complexity, in terms of the reorganization of elements.

4.11.4 Relative Complexity

WH-Structures

WH-Questions with Inversion

Viewed in the context of its corresponding declarative structure, the first example cited involves a reorganization of the peripheral elements of subject and complement around the verb: *SUBJ V COMP* becomes *WH_WORD V SUB*. In transformational terms, this involves two movement rules: a preposing rule where the complement moves from final to initial position, where it is realized as some form of WH-word, and a rule of subject-verb inversion, which leaves the subject in final position. It is not clear, however, what sort of psychological reality such transformational descriptions have, especially as descriptions of trajectories of acquisition [76]. Structures of the kind which should result after the application of the first movement rule, namely *WH_WORD SUBJ V*, do occur, but are not especially frequent in the present data:

'Partontatis', what it is?

- kb.1: [652] -

What you can do for your country?

- canh.1: [2432–3] -

One problem with looking at the complexity of learning and processing tasks in transformational terms is that in learners' grammars, as is abundantly clear, a particular structure may not be analyzed into the same constituents as in the target grammar. And with fewer elements in a structure, the degree of complexity may very possibly be less.

This is a consideration to be taken into account when comparing the two types of WH-question, since it may well be the case (and there is certainly plenty of analogous evidence from areas such as pronominalization) that the form for WH-words is erroneously perceived in the first instance by the learner to be what is in fact a combination of verb and WH-word. That is, forms such as what's, where is and how are may initially be monomorphemic in the learner's grammar. The rarity of *WH_WORD SUBJ VERB* structures (of which examples were just given) when the verb is the copula can be explained as a consequence of the target pattern of *WH_WORD COPULA* being well established through early bound forms. Other support for the initial absence of a separate copula in early learner language comes from its frequent non-appearance in the declarative forms which parallel *WH/INV* questions: if these equational sentences can consist of only two elements, then related interrogative forms probably can too. Given all this, then,

only one movement rule might be required to produce *WH/INV* structures like What is this?—that is, at least until such time as decomposition of the WH-form takes place.

WH-Questions with 'Do'-Insertion

Similar considerations apply to the second WH-structure. If all of its elements are considered to be separate, then the final structure results from a rule preposing the object, and a second rule governing the insertion of do (or one of its derivations) between the WH-word and the subject. In processing and transformational terms, this structure could be presumed to be more complex than the first, in that both rearrangement and the introduction of new material have to take place. If, on the other hand, do is initially bound to the WH-element, only one preposing rule is required, and the level of complexity, in the first instance, would be similar to that of the previous structure. However, once the learner begins to register that do is not in fact bound to the WH-word, the situation as regards complexity changes. Decomposition of WH-questions with inversion into three principal components would reveal a close parallel with affirmative equational structures, and would presumably be actually associated with the stabilization of the rule system for these structures. The structure would find support in related paradigms. Decomposition of WH-questions with do, on the other hand, would leave the learner with the task of having to account for do, a morpheme whose pattern of distribution elsewhere would offer few clues at this point (remember that in the data on negation it is always bound) as to its probable function. In addition, the form-function constraint could impede any assignment of function to do, given that a lexical verb with the same form would have almost always been acquired by this time.

WH-Structures—Summary

The above considerations, then, might suggest that predictions of learning difficulty on the basis of transformational complexity need to take into account the fact that the structural descriptions for a particular unit may not be the same for native speaker and learner. At this stage, as well as processing considerations the learner has the problem of segmentation of constituents to deal with. In the process of segmentation in question structures, the learner will be faced with the problem of assigning a place in his or her grammar to elements which turn out to have separate identities, such as do.

This approach to the effect of transformational and processing complexity on learning difficulty takes into consideration both the important principle that learning involves analysis and segmentation, and could perhaps offer a more "psychologically real" account of how transformational com-

plexity might actually manifest itself in a dynamic system such as a learner grammar. Clearly, the suggestions that have been made above are tentative. The author is aware that effective predictions of difficulty for word order rules in German have been made on the basis of clearly defined processing constraints [34]. The same kind of enterprise needs to be undertaken for English. A problem with direct application of the German model is that it is designed to operate on strings with a constant number of elements. In English, phenomena like *do*-insertion which introduce lexical morphemes into the string to be processed have to be dealt with as well. It is not clear that this can be accomplished using a purely processing model—that is, processing considerations may have to supplemented by syntactic ones, since differing derivational histories for newly introduced elements may have different computational costs [128].

Taking into account that constituents in learner grammar and the target grammar may be different, for the two WH-structures it could be proposed that initially, if and while the WH-word remained bound to the verb, they are more or less equally complex, and, therefore, likely to exhibit similar, restricted patterns of use. At the point when their internal structures began to emerge more clearly it could be expected that the structure with do would present the learner with more problems than its counterpart, and would therefore spread more slowly.

Yes/No Structures

The situation is somewhat more straightforward for the first of the Yes/No structures. While binding of the verb to the subject cannot be ruled out, the wider distribution of pronouns and noun phrases renders it far less likely, and in any case it is difficult to see what would motivate such a phenomenon. Therefore, a simple inversion transformation seems the most likely path of formation for this structure.

In the case of the second Yes/No structure, a rule of *do*-insertion seems totally unavoidable, in the absence of any viable bearer for the morpheme, as is the case in WH-questions. This means that a learner who consistently produces such structures will have to accommodate do into his or her grammar in some capacity. If it does not function as an auxiliary, at the very least it would have to have the status of a question marker. The learning task presented by this structure appears to be the most complex of the four described: new material has to be inserted into the structure and in the absence of any convenient bearer, such as the WH-words provide, this new material has to allotted some definite function in the learner's grammar. In contrast to the case of *do*-insertion in WH-questions, this problem is evident at the very outset. Once again, the form-function constraint may complicate matters still further.

4.11.5 Apparent Learner Difficulty with Question Structures

Having put forward some hypotheses, it is time to turn to the data. The first point to note is that the majority of questions produced by informants at all stages of development, and presumably native speakers as well, are structurally incomplete phrases marked by intonation.

Er...er, any?

- jb.1: [39] -

Zygmunt?

- jr.1: [166] -

Wha(t)?

- duc.2: [181] -

In Poland?

- jb.2: [352] -

These forms serve to confirm and repair and generally facilitate the flow of discourse. In many cases their elliptical nature would be quite acceptable in target language terms; in some it would not. With regard to an analysis of syntactic development they do not provide much useful evidence, but it is certainly worth remembering their importance for the communicative system as a whole.

Of the structurally more complete tokens of interrogation in the present study, a great many have been discarded as formulaic. Typical examples of formulaic questions are the following:

Where do you work?

- bb.1: [1102] -

What is it?

- duc.1: [379] -

Are you marri(ed)?

- dung.1: [694] -

How ol(d) are you?

- dung.1: [712] -

would you li(ke) some cake?

- hoa.1: [527] -

do you know Civi(c) Hou(se)?

- hoa.2: [219] -

What's your name?

- jb.1: [911] -

Apart from the unequivocal cases, there is a further group of borderline cases, where, for instance, an apparently creative application of the rules in a particular informant's system turns out to be restricted to, say, a single verb. The procedure adopted in the present study is to look at these cases in relation to the non-standard forms produced by an informant. Where this latter class is extensive, and provides little supporting evidence of a system in which rules are developing, there are strong grounds for treating the borderline cases as formulas, rather than evolving forms. Tables 4.41 and 4.42 provide counts for the number of complex but non-standard structures produced by each informant.

For the remaining tokens, the patterns, for reasons which will become clear, differ somewhat for the two language groups. Tables 4.41 and 4.42 show the distribution of the four main question structures for each language group. Due to the relatively small amounts of data involved, the two interviews were treated as a single source; this is not, however, the normal practice in this study.

The notational conventions for the Tables are as follows:

1. A trailing *F* means all the tokens in a particular count were judged to be formulaic.
2. A trailing *F?* means that the tokens may be formulaic.
3. A trailing *X* indicates that the token is non-standard in some way, but still a valid example of the structure.

Polish Speakers

For the Polish speakers, the most widely distributed and frequent question type is the WH-question with inversion. The majority of these tokens were judged to be formulaic. Informants who produced almost certainly non-formulaic tokens are Krystyna A., Barbara B., and Ludwiga J.

how long can we... er... live here in hostel?

- ka.1: [1133] -

and where is the winter?

- bb.1: [494] -

what was happened?

- lj.1: [141–2] -

The token produced by this last informant, illustrates a paradoxical quirk of evaluating data for likely formularity: non-standard tokens, because they cannot have been heard and learnt as single chunks (at least from native speakers), can provide the most convincing examples of the existence and application of a set of underlying rules for a structure.

The situation for WH-questions with inversion is in accordance with the hypotheses advanced above. That is, the form, possibly at the outset with a less complex internal structure, seems relatively easy to learn and use. The production of non-formulaic tokens by learners from the middle ranges of the ASLPR onwards suggests that the transition from formula to rule based structure is comparatively straightforward.

The next most widely distributed question type is the Yes/No question with inversion. Examples are:

Are you a teacher?

- es.1: [1184] -

have you ever children?

- ka.1: [1200–1] -

can you tell me, please?

- kb.1: [1093] -

The distribution for this question type is similar to that of the *WH/INV* type. This is in accordance with the predictions made about the complexity of the two structures.

WH/DO questions are the next most widely distributed structure. Of the tokens produced, however, all but one are either formulaic or possibly formulaic. Some examples are:

when do you... come... er... next?

- es.1: [1146] -

where do you working?

- ks.1: [947] -

The one exception is somewhat unidiomatic, which as it turns out is not uncommon for this structure:

what do you think about my future..here?

- kb.1: [1083–4] -

In the case of the fourth structure, Yes/No questions with do, it is not certain that there are any non-formulaic tokens in the data. The examples from Krystof S. and Jerzy B. are:

do you married?

- ks.1: [962] -

Er, do you want, ah...?

- jb.2: [576] -

In the case of Krystof S., do could possibly be misclassified as a question marker—if so this would provide an interesting insight into how an early learner grammar attempts to deal with this form. Alternatively, given the token of do you working from the same informant cited above, this learner may have only one auxiliary. This, too, is interesting evidence, albeit too isolated for any conclusions to be drawn. The other case, lacking a complement, must remain doubtful.

Krystyna B., produces six tokens of *YN/DO* structures. However, five of these involve the likely formula do you understand...?, and the sixth the related do you know?:

do you understand me what I say?

- kb.1: [862] -

The lack of tokens of any kind for *YN/DO* structures, and the obvious difficulties in formulating a set of working hypotheses to enable the structure to be generated are exactly what was predicted. As discussed above, learners need some place in their grammar for do right from the outset to produce these structures. Both from the evidence of questions above and negation, this slot is a difficult one to provide.

Vietnamese Speakers

Turning now to Table 4.42, we find that for the Vietnamese speakers *WH/INV* is the most widely distributed question type.

What is it?

- duc.1: [390] -

How many children have you got?

- dung.1: [698] -

How o(ld) are you?

- hoa.1: [402] -

In contrast to the case of the Polish speakers, however, there are no clearly non-formulaic tokens in the data.

The reason for this appears to lie in the low rate of copula insertion amongst Vietnamese speakers. Thus, while the Vietnamese learners can acquire and produce *WH/INV* structures in which the WH-word and the copula form an unanalyzed unit, their lack of sensitivity to copular structures in general means either that they experience difficulty in parsing these structures so that they can subsequently be productively generated or that they simply produce *WH/INV* structures minus the copula. There are indeed examples of the latter in the data:

How many people in your family?

- dung.1: [720] -

what your nationality?

- minh.1: [429] -

whas your name?

- my.1: [8] -

Of course, the copula is not the only verb which can appear in *WH/INV* structures, though it is probably the predominant one in learner speech—certainly from the evidence from the Polish speakers and the formulaic tokens.

The low rate of copula insertion for the Vietnamese group no doubt affects the distribution and frequency patterns for the other question type in which inversion figures, Yes/No questions. There are in fact a number of possibly non-formulaic tokens of this structure. All of them, however, involve verbs other than be:

would you tell me

canh.1: [1814–5]

Have you got a house?

- dung.1: [773] -

can you as(k) me again?

- hoa.1: [655] -

Question formation, given that it involves ordering rules and identifiable processing constraints, is a likely developmental feature. It is therefore of some significance that such situations as this one with inversion in the speech of Vietnamese learners can occur, where the rules for a variational feature like copula insertion interact with and inhibit the rules for a developmental one.

As a result of the constraints on inverted structures for the Vietnamese learners, it is *WH/DO* structures which exhibit the widest distribution of non-formulaic tokens. Examples of these are:

Where did you come from?

- dung.1: [716] -

How much did you buy your car?

- dung.1: [800] -

how many do you have, eh... children?

- minh.1: [444] -

what do you mean 'problems'?

- phuc.1: [656] -

As in the case of the Polish speakers, some of these tokens are non-standard in other regards. A common problem is choice of the wrong WH-word:

what does, ah, Fren(ch) call 'sans b'argent'?

- canh.1: [1742] -

How do you thin(k)?

- tam.1: [1455] -

This highlights another difficulty—semantic rather than syntactic—posed by this structure but not by its *WH/INV* counterpart. Given a much wider possible range of verbs for *WH/DO* questions, the likelihood of a mismatch occurring between verb and WH-word is far greater.

For the last of the four question types, Yes/No questions with *do*-insertion, the results are much the same as for the Polish speakers. The number of informants who produce any kind of token of this structure is less than for the other three types. Only one informant, Sang, produces probably non-formulaic tokens of *YN/DO* questions. Three of the four tokens are non-standard, in that did is used when do would be the standard morpheme. The reasons for this are not clear, although it could be an attempt to mark modality (cf. would), or else a way of avoiding the duplication of lexical do.

Did you li(ke) to have a f(r)ien(d) Vietname(se)?

- sang.1: [598] -

Did you li(ke), er, to hear abou(t) my story?

- sang.1: [692–3] -

4.11.6 The Order of Difficulty—Summary

In summary, the order of difficulty for non-formulaic question types for the Vietnamese speakers is not the same as for the Polish speakers, whose results were as predicted by the hypotheses put forward at the beginning of this section. The reason for this difference lies in a different pattern of copula insertion for the Vietnamese learners, which affects the production of structures involving inversion. For the structures involving *do*-insertion, the pattern is as predicted, with the most difficult structure, *YN/DO*, once again hardly in evidence.

It should be pointed out that the pattern of distribution for formulaic or partially analyzed structures differs only slightly from that for the Polish speakers.

The case of *WH/INV* structures in the Vietnamese group is a very good example of the necessity of distinguishing formulas from rule based applications—a more cursory approach to the data would have failed to isolate an important difficulty for this group of learners.

As far as both groups of learners are concerned, and always keeping in mind the limitations of the data itself, it seems reasonable to say that the

hypothesized difficulty of the four structures appears to be reflected in the above findings. More definite confirmation will have to wait until a more extensive corpus of data on questions from a wider range of situations can be collected and analyzed. An analysis of the elicited data, where the question type is largely dictated by the form of the prompt, and avoidance patterns should therefore be obvious, would provide another means of examining the relative difficulty of these four structures, and is an evident direction for further investigation.

4.11.7 Ill-Formed Structures

To this point, the discussion has been restricted to well formed, or apparently well formed structures. The final column in Tables 4.41 and 4.42 shows the figures for structures of more than phrasal complexity which failed in some way to meet the criteria for well-formedness. As it turns out, there are numerous ways in which this can happen.

The most straightforward cases of non-standardness are those in which either inversion or *do*-insertion failed to take place. Examples of the non-application of inversion are:

where I can got, er... information?

- aj.1: [774] -

how old... he is?

- jr.1: [1322] -

They are, er, Vietname(se) maps?

- canh.1: [68] -

you can get job like me now or not?

- long.1: [785–6] -

It was difficult to find many examples of WH-questions where inversion had not been applied in the output of the Vietnamese speakers due to the strong tendency of such structures to lack a copula at all.

Examples of the non-insertion of do are:

how you ca(ll) de name?

- hoa.2: [189] -

What you want?

- long.1: [621] -

you live in Australia... yes?

- jr.1: [1286] -

You b(e)lie(ve) in da(t)?

- sang.2: [687] -

Many other types of non-standardness result from attempts to avoid the word order problems presented by questions.

One such strategy is non-insertion of the subject:

not... er... miner... er... what is?

- bb.1: [60] -

Where is?

- ka.2: [355] -

and... is good profession?

- bb.1: [1130] -

Is, er, near city?

- jr.1: [1268] -

In addition to the subject, other elements can also fail to appear:

How long... from Poland to Australia?

- ka.1: [116] -

understand me?

- bb.2: [316] -

no(t)s right?

- vinh.2: [419] -

Another strategy involves leaving out the copula, as in these examples:

whas your name?

- van.1: [119] -

what you... hobby?

- jr.1: [1338] -

This all?

- bb.1: [1049] -

Your sore? [sure]

- hoa.1: [457] -

Or an auxiliary of some kind:

How long you (d)rive car?

- minh.2: [271–2] -

How long I waiting?

- jr.1: [410] -

You think so?

- tam.1: [1357] -

Excuse me, I go to little while?

- canh.1: [2447] -

There are other possibilities as well, such as not preposing the WH-constituent:

Have you got, eh... how many children?

- duc.1: [744] -

It should also be mentioned that certain types of question are conspicuous by their total (or almost total) absence. The most notable examples are negative questions and tag questions. The range of verbs on which questions are formed is also quite restricted. The examples throughout this section provide a representative sample of these.

4.11.8 Embedded Questions

One other phenomenon for which some results exist is the production of embedded questions—that is, questions which are actually the complement of some verb like know or ask:

I don't know what I must said

- kb.1: [97] -

I don('t) care where... where I live

- phuc.2: [135] -

Embedded questions are probably a fairly good index of overall structural sophistication. One interesting characteristic of embedded questions is that the rules of inversion and *do*-insertion do not apply in them, and may therefore have to be unlearnt by learners who acquire them. This of course depends on whether they have really learnt such rules in the first place. Tables 4.42 and 4.43 provide a list of those speakers who produce embedded questions and notes whether movement or insertion rules have applied. To evaluate the significance of this latter information, it is necessary to refer back to Tables 4.41 and 4.42 for evidence of what rules a particular informant applies in direct questions. Embedded questions can, of course, take various other forms—such as infinitival or intransitive—where movement or insertion rules are not applicable, or they can be ill-formed. No attempt has been made to detail or quantify such features in the tables.

4.11.9 Conclusions

There are various points of significance to emerge from the analysis of questions.

There appears to be a definite hierarchy of difficulty amongst the main question types, with forms involving inversion less difficult than those involving *do*-insertion.

The introduction of new material into a structure—or its discovery as the structure is analyzed—seems to present a more complex processing task than movement of existing constituents. One of the difficulties with new elements is that learners must adjust their grammars to accommodate them, either as new items or as old items with new functions.

Do is a particularly difficult case of such a new element, given its prior function as a lexical verb. Yes/No questions with *do*-insertion, where there is no possible way for do to enter the system as part of another element are the most difficult of the four main question types. WH-questions with do, while easier to acquire in partially analyzed form, with do as a part of the

WH-word, present problems to the learner once he or she begins to try to analyze and generalize them. With this question type there are semantic complications as well arising from the necessity of matching the WH-forms with a wide range of verbs. The state of a learner's system with regard to copula deletion can affect structures involving inversion, as it does with the Vietnamese speakers.

The basic sequence itself, and factors such as these need to be taken into account when teaching interrogation. In general, the range of verbs used in questions is quite limited, and this too should be considered by the teacher. Some forms, such as negative questions and question tags should probably be left for a later stage. Embedded questions, on the other hand, are produced and could be taught; they represent a communicatively useful way of building up more complex structures. A fair percentage of non-standard question structures characterize the output of most informants in the sample: this should probably be tolerated by the teacher. It is worth remembering that the movement and insertion rules in questions are not usually crucial to communication, provided that there is some kind of intonational flag. Suppression of major constituents like subjects or verbs tends to reduce the communicative effectiveness of questions more than absence of inversion or *do*-insertion and corrective work might be more fruitfully focussed on these features; at the same time the teacher should bear in mind that phenomena such as subject deletion may be evidence of a change in the learner's rule system, and should be temporarily tolerated.

4.12 Nominal Morphology

4.12.1 Inflectional Morphology

This section deals only with inflectional nominal morphology, namely regular plural *-s* and genitive *-'s*.

Previous studies of the "acquisition" or order of difficulty for these morphemes, such as morpheme order studies, have assumed that the two morphemes in question were developmental features, acquired in a fixed order in relation to other morphemes. As we shall see, in the case of plural *-s* at least there is some cause for doubting this.

4.12.2 Plural '-s'

During the transcription process in any context where the plural form appeared to be obligatory and was not realized this was flagged by the addition of either (s) or (es), as appropriate, to the word in question. This practice of using round brackets to signal non-realized morphemes or phonemes is, as noted in 3.3.4, a general convention of the transcriptions. Examples of null plural marking flagged in this way would be:

Informant	WH/INV	WH/DO	YN/INV	YN/DO	NONSTD
Irena S.					
Zygmunt J.	1F				1
Krystof S.	5F	2F?	1F	1X?	5
Jerzy B.	4F	1F	1F	1F?	5
~	1F?				
Jan R.	3F				27
Krystyna A.	2F		2X		5
~	2				
Barbara B.	5F	5F	1		23
~	1				
Ewa S.	1	1F?			4
Andrzej J.			1F?		4
Krystyna B.	2F	1X	4	6F?	12
Ludwiga J.	6F	1F	2F		10
~	1X				
Mieczyslaw M.			1X		6

Table 4.41: Distribution of Question Types: Polish Informants

Informant	WH/INV	WH/DO	YN/INV	YN/DO	NONSTD
Van	2F				2
My	2F				5
Duc	3F				9
Dung	2F?	8	1F	1F	
~		1F?	3F?		2
Minh		1F	3	2F	7
~		1X			
Hoa	1F	1F?	3F?	1F	8
Sang		2	2F	4X	3
Vinh					3
Tam	1F	1X			12
Canh	1F	1F?	3	2F?	20
Long			1F		16
Phuc		1X			2

Table 4.42: Distribution of Question Types: Vietnamese Informants

Informant	Emb_Qs	Do/Inv	No_App
Krystof S.	1		
Barbara B.	8	X	
Ewa S.	2		
Andrzej J.	9	X	X
Krystyna B.	12	X	X
Ludwiga J.	13	X	X
Mieczyslaw M.	3	X	X

Table 4.43: Distribution of Embedded Questions: Polish Informants

Informant	Emb_Qs	Do/Inv	No_App
Dung	3	X	
Minh	3		
Sang	3	X	X
Vinh	2		
Tam	5		X
Canh	10	X	X
Long	12		X
Phuc	14	X	X

Table 4.44: Distribution of Embedded Questions: Vietnamese Informants

I learn two... er... job(s)

- bb.1: [44] -

Yeah.../... go to parti(es)

- dung.1: [122] -

Close attention was paid to this procedure during the checking of the transcriptions, so that the counts thus obtained should be reasonably accurate.

Methodological Problems

In checking for plural marking there are various complicating factors, and it is probably worth noting these.

1. It is not always clear that a particular semantic context is obligatorily plural. For example, consider the following exchange:

   I Em... get... Poland, er... my, er... my, er... document... Poland document mus(t) ... mus(t), er... travel?... no

   MJ Translate?

   I Translation, yes

   - ks.1: [124–9]

1. It only becomes clear further on in the interview that several documents are involved, so there is no flagging of the null plural at this point.

2. There can also be problems with the classification of generics, where it is difficult to decide if it is a determiner or the plural morpheme that has not been supplied, as in the following example:

   and their hands are on trigger... trigger ready to shoo(t)

   - phuc.1: [797–8] -

1. There are also some problems of phonetic origin. Vietnamese speakers frequently produce an */s/* phone, or something that sounds very like it to the ears of an English speaker, in place of a target alveolar stop. Thus but becomes bus, or good, gooz. This means that it sometimes difficult to know whether to credit a speaker with plural marking or not.

2. Another case, this time involving both groups, has to do with the word months, which is a frequent candidate for pluralization. Here the complication is that the singular is frequently rendered as mons (*/mans/*), and there is really no way to tell whether an informant is attempting to pluralize a token unless he or she produces both the */th/* and */s/* phonemes, which is something that even native speakers may neglect to do in casual speech. (A check on this latter phenomenon revealed that omission of tokens of months did not significantly affect any of the counts).

3. Moreover, in a more detailed analysis of plural marking, it would probably be necessary to take into account the contexts in which marking does or does not occur. There are various possible scenarios. A learner who marks plurals with some regularity in noun phrases which follow plural numbers or quantifiers, or in other contexts where plurality is evident, may not know the rules, say, for generic noun phrases. Another possibility is that an informant might mark plurals more regularly where such marking is not redundant, so that he or she may actually be less likely to mark noun phrases after plural numbers or quantifiers than elsewhere. In the present analysis only the apparently most congenial lexical environments for plural marking are investigated.

Distribution of Plural '-s'

Bearing these considerations in mind, Table 4.45 provides some figures on the number of realized and null plural morphemes for each informant.

The most obvious point in the findings is the discrepancy between the Polish speakers and the Vietnamese speakers. Amongst the Polish speakers there are only two cases of null tokens exceeding realized ones—**bb.1** and **jr.1**. Amongst the Vietnamese speakers, on the other hand, there are only four cases in which realized tokens exceed null ones—**phuc.1**, **phuc.2**, **minh.2** and **dung.2**. Unquestionably phonetic and phonological difficulties contribute heavily to the state of affairs with the Vietnamese speakers. Other factors will be discussed below.

The above result naturally calls into question the status of the plural morpheme as a developmental feature, given that it appears to exhibit extreme vulnerability to first language influence. This doubt is magnified by the results gathered from the Polish speakers, in which some of the unquestionably more developmentally advanced informants have lower ratios of realized to null tokens than the less advanced learners. (Compare the results for Irena S. and Ewa S., for instance). In addition, while there is a general improvement in the ratio of realized to null tokens in the second interview for the Polish speakers, there are some apparent cases of "backsliding", such as Krystyna A. and Krystyna B. This is not what one would

WORDS	INTERVIEWS
years	36
months	20
sometimes	18
friends	16
hours	16
weeks	16
days	15
times	14
things	13
books	12
dollars	11
problems	11

expect of a developmental feature [128].

In 4.1 a rough characterization of "standardizing" versus "simplifying" learner types was presented (Table 4.7). For the Polish informants, while the correspondences between learner type and rate of suppliance of plural *-s* is by no means regular, there seems to be some correlation between these two characteristics. This correlation might possibly improve if learner type was assessed by some less intuitive means.

In the case of the Vietnamese speakers, their phonological problems with final stops and consonant clusters complicate the issue to such a degree that there must be some doubt about the value of discussing plural *-s* for these speakers in either variational or developmental terms. Despite the gap between them and the Polish speakers, however, it should be remarked that once again in general the standardizing learners appear to have a significantly higher rate of suppliance than the simplifying ones.

Clearly, the status of the plural morpheme requires further research, with a larger corpus and a wider range of language backgrounds. On the basis of the present data, however, grave doubts about the developmental nature of this particular morpheme are unavoidable.

Lexical Candidates for Pluralization

While some 275 different words are pluralized across the whole corpus, a relatively small subset of these are pluralized repeatedly. Words that are pluralized in more than ten of the forty-eight interviews in the cross-sectional sample are the following:

As regards phonetics, the unpluralized stems of these words (as they are actually pronounced by learners) terminate predominantly in nasal consonants (six cases), vowels (four cases), and voiceless velar obstruents (two

cases). The nasal consonants may in some cases be replaced by a nasalized vowel, which would have the effect of avoiding a final cluster and preserving a vowel-consonant order. From a semantic point of view, seven of the words are time-related. While the list given above no doubt reflects the discourse parameters of the interview situation, it would be advisable to keep its composition in mind for the development of explicit materials for the practice of regular pluralization. If the phonetic observations made above are valid, they would, of course, be largely independent of particular choice of words. This topic could be fruitfully researched in more detail. Tables 4.50 and 4.51 provide figures for the distribution of the most commonly pluralized words in the first round of interviews.

Overgeneralization of Plural '-s'

As with other morphemes, plural *-s* is prone to a certain amount of overgeneralization or semantically inappropriate application. Some of this is idiosyncratic, as in the following examples:

I remember one girls they workin(g) together

- long.1: [723] -

Doctors?... yes... er... my sister is doctors

- bb.1: [624] -

Other examples, however, are more widespread, and call into question the notion of pluralization with some words:

and one years is not problem

- bb.1: [958] -

and one hours

- hoa.1: [39] -

one years... two years

- is.2: [639] -

Other examples show that there has been a failure to recognize irregular or partitive plurals:

I fishermans

- long.1: [139] -

I must to understand many peoples

- lj.1: [60] -

can I have a pants please

- kb.1: [689] -

oh, some peoples maybe

- kb.1: [834] -

The Plural Morpheme—Summary

Although further research is required, the developmental status of the plural morpheme is very dubious. This morpheme seems to be particularly susceptible to phonetic (first language) interference, and appears to display other characteristics of a variational nature. Where it does occur, certain phonetic environments seem to favour its suppliance. In addition, in the present corpus at least, there is a group of words which are likely candidates for pluralization. These factors have obvious pedagogic significance.

4.12.3 Possessive "-'s"

Possession can be lexically, syntactically, or morphologically indicated in Standard English. Lexically indicated possession involves verbs such as have and belong and involves structures in which the possessor may be grammatically a subject or an indirect object, with the choice of grammatical relations being determined by sentence pragmatic or discourse considerations. Syntactically indicated possession involves possessed-possessor structures linked by the preposition of and is subject to various semantic constraints, which restrict its usage quite powerfully. Morphologically indicated possession involves case marking of the possessor noun phrase with the ubiquitous *-'s* morpheme, and a word order in which possessor precedes possessed. These latter structures can be more generally implemented than possessed-possessor structures, though they are by no means unconstrained, and are what we generally think of as possessives in English. As a result of their dependence on the *-'s* morpheme for case marking of the possessor, they have, through the years, attracted a good deal of attention from the proponents of morpheme order acquisition studies—where the general consensus is that the possessive *-'s* morpheme is the "most difficult" or last acquired of the three *-s* morphemes [11].

Interview	Tokens
aj.2	4
phuc2	4
mm.2	2
duc.1	1
dung.1	1
is.2	1
kb.1	1
phuc.1	1
es.2	1F
vinh.2	1F
canh.1	1?
es.1	1?

Table 4.45: Distribution of Possessive "'S"

In the present study, the evidence for this is somewhat equivocal. On the one hand, the frequency and distribution of the three *-s* morphemes is in line with the predictions of morpheme order studies. However, as we shall see, closer examination of the data reveals violations of the predicted implicational order and a general lack of developmental patterning for the nominal inflections.

Distribution of Possessive "-'s"

Tables 4.45, 4.46, 4.48 and 4.48 show the distribution of possessive *-'s*.

At this point, no figures on obligatory contexts are available for comparison with those given in Tables 4.45, 4.46, 4.48 and 4.48. Figures for the realized morpheme are not high. Tokens of possessive *-'s* occur in only twelve of the forty-eight interviews in the cross-sectional study. In two interviews the tokens are formulaic, that is, they involve names—Regent's Park and Mario's Drivin(g) School. In two others they are dubious for semantic reasons. This reduces the number of certain cases to eight. The maximum number of tokens produced in any one interview is four. Excluding the dubious or formulaic tokens, the number of informants who produce tokens of possessive *-'s* is seven.

Alternative Means of Indicating Possession

Clearly, then, possessive structures do not constitute an important part of even relatively developed learner language.

For learners, alternative ways of indicating possessor-possessed relations are available. One strategy is to simply juxtapose the possessor and the

possessed in an noun phrase without producing an morphological marker. Examples are:

the gir(l) name Dung too

- hoa.1: [952] -

Possessed-possessor structures with of also occur:

I've look... looked at, er... windows of the travels agency

- kb.1: [306–7] -

Non-standard possessed-possessor structures minus the preposition also occur:

husband my, eh, shvester.../...cou(ld) found job

- ks.1: [556] -

Other more periphrastic constructions may also occur occasionally:

belong car his brother-in-law

- dung.1: [659] -

4.12.4 The Developmental Status of '-s' Morphemes

As with the plural morpheme, the data on the production of possessive *-'s* raises questions about the developmental status of this morpheme.

In the present data, while the general pattern of *-s* morpheme production coincides with the findings of morpheme order acquisition studies, closer examination of the data on the nominal *-s* morpheme and comparisons with other kinds of data does not reveal any internal structure—in the nominal *-s* morpheme data at least—which would suggest that production of the *-s* morpheme occurs in a developmentally regular or coherent way.

This lack of developmental regularity has already been discussed in some detail in respect of plurals. If we examine the behaviour of those informants who produced tokens of possessive *-'s* in relation to the other two *-'s* morphemes we find that the seven informants who produce tokens of possessive *-'s*, fall into two groups. Four—Phuc, Mieczyslaw M., Krystyna B., and Andrzej J.—rank highly on the *ASLPR*, exhibit generally advanced structural and morphological development and relatively high production figures for third singular *-s*. (Though "relatively high" is from all rough indicators likely to be "rather low"). On the other hand, these informants' control of

plural marking is somewhat variable. Thus, while Andrzej J. and Mieczyslaw M. have, respectively, relatively high and high ratios of realized to null plurals (see Tables 4.45, 4.46, 4.47 4.48 and 4.48), Krystyna B. and Phuc exhibit lower ratios. Despite this, we might reasonably expect some tokens of possessive *-'s* from the informants in this group from within a morpheme order framework.

Members of the second group—Duc, Dung, and Irena S.—rank low on the *ASLPR*, however, and generally do not exhibit the structural and morphological development of the first group of informants. Irena S. has middling ratios for realized versus null tokens of the plural but produces no tokens of third singular *-s*. Duc and Dung have lower ratios for plurals but produce a few valid tokens of third singular *-s*, although Duc in fact produces this morpheme three times out of four with first person subjects and always in past tense contexts.

If there was a clear implicational relationship of the $POSSESSIVE \supset THIRD\ SINGULAR \supset PLURAL$ type, we would not expect possessive *-'s* tokens from Irena S. or, probably, Duc.

4.12.5 An Alternative Proposition

In trying to make sense of this anomaly we should note that while the informants in the second group rate less well than those in the first on plural *-s*, what really distinguishes them is their lack of control of the third person singular morpheme. This suggests that in fact the three *-s* morphemes should perhaps not really be ordered as they are on a single implicational scale.

From a speech-processing point of view, it may be possible to make out a case for the above assertion. Roughly, this is as follows. In nominal morphology the governor for marking is within the same noun phrase complex as the marker—either in the form of a number or quantifier, or inherent in the semantics of the utterance. In verbal morphology—in the case of third person singular marking at any rate—the governor is another constituent—namely, a pronoun or noun phrase and the marker is separated from its governor by the verb [93]. Thus marking involves sentence-internal rearrangement of material. In processing terms this type of interruption leads to greater complexity. Non-standard structures in which this internal rearrangement has not occurred can indeed be found in learner speech:

Yes... yes... he's have

- ka.1: [624] -

he's tell me come

- long.1: [173] -

she's remember me and my sister

- vinh.1: [348] -

The upshot of the above is that it is possible that of the three *-s* morphemes, the third person singular is in fact a developmental feature, while the plural and genitive are not. This explanation fits the present data more satisfactorily than the traditional implicational ordering. Clearly, with so little data for the genitive it is hard to determine its status *vis-a-vis* the developmental/variational distinction; in the case of the plural we have seen evidence that it is very possibly not a developmental feature. From a speech-processing point of view, there seems even less reason to assume that possessive *-'s* is a developmental feature, since the marking entailed involves no reorganization of constituents.

The leaves us, of course, at loggerheads with the findings from morpheme order studies. It should be noted, however, that a critique of these studies has been offered in 2.1.5 which raises serious doubts as to the validity of their findings. The low figures for the production of genitives there and in the present data may, as has been suggested reflect nothing more than the relative frequency of this item in the typical range of discourse situations [109]. As with other varieties of the *-s* morpheme, phonetic considerations and the form-function constraint may contribute to limitations and irregularities in the suppliance of the genitive. Once again, further research will have to be conducted if these issues are to be clarified.

4.12.6 Nominal Inflections—Summary

In regard to nominal inflectional morphology, it is possible that the traditional morpheme order study order of difficulty or acquisition is erroneous, given that there exist substantial doubts about the developmental status of these markers. There is some evidence that regular plural marking may be a variational feature. Due to lack of data the status of the genitive is unclear, but, equally, it may not be developmental in nature.

Regardless of this, there is evidence of some phonetic and possibly lexical regularities in the acquisition of pluralization. These should be noted. Likewise, there is evidence that possessive *-'s* is barely used by learners in the proficiency ranges covered by the study, and consequently it should not be allocated a very high pedagogic priority.

4.13 The Definite and Indefinite Articles

4.13.1 Limitations

Tables 4.52, 4.53, 4.54 and 4.55 to 4.13.1 provide some basic information about definite and indefinite article usage in the cross-sectional sample.

Informant	is	zj	ks	jb	jr	ka	bb	es	aj	kb	lj	mm
Realized	12	20	21	18	17	27	16	27	34	47	99	55
Null	6	10	15	16	37	7	38	25	11	17	39	4

Table 4.46: Plural "-s" Morpheme: Polish Informants—1

Informant	is	ks	zj	jr	ka	jb	bb	es	aj	kb	lj	mm
Realized	21	23	30	13	10	34	21	17	52	28	82	60
Null	3	22	7	15	6	7	16	15	9	25	43	7

Table 4.47: Distribution of the Plural "-s" Morpheme: Polish Informants—2

Informant	Van	My	Duc	Dung	Minh	Hoa	Sang	Vinh	Tam	Canh	Long	Phuc
Realized	1	1	20	18	4	3	1	9	14	44	33	89
Null	6	6	34	18	20	32	23	41	57	121	108	29

Table 4.48: Plural "-s" Morpheme: Vietnamese Informants—1

Informant	Van	Duc	Sang	Minh	My	Hoa	Dung	Vinh	Tam	Canh	Long	Phuc
Realized	-	8	7	7	2	-	19	1	16	6	8	18
Null	7	27	28	5	43	25	13	30	41	40	10	9

Table 4.49: Distribution of the Plural "-s" Morpheme: Vietnamese Informants—2

Informant	is	zj	ks	jb	jr	ka	bb	es	aj	kb	lj	mm
years		5	8	5	4	8	6	2	10	10	16	7
books	2			2	1	1		2	1		2	2
sometimes		1		1	3	1			2	1	9	1
months			4		2	3		1	1		5	
hours				1	1	1		1	3	1		
weeks			1	1					4	2	3	
friends		4						3	2		5	
times			1						2		4	2
things				1						1	2	6
days	1				1			2				
problems	1						1				6	
dollars							4	1				

Table 4.50: Most Frequently Pluralized Words: Polish Informants—1

Informant	Van	My	Duc	Dung	Minh	Hoa	Sang	Vinh	Tam	Canh	Long	Phuc
years			1	4	1			6	6	5	14	8
months			1	5	2			1		1		1
days		3	3	1			1			2	4	
hours		1		4		2						3
friends						1			4		1	1
times			1							1		1
weeks					1		1					1
things								1		4		1
dollars				3							4	
sometimes						1						4
problems												2

Table 4.51: Most Frequently Pluralized Words: Vietnamese Informants—1

Table 4.13.1 gives the absolute number of tokens of the, a and an, or their phonological variants for each interview. In the case of the first-round interviews these tokens were hand checked and repetitions or erroneous tokens eliminated. This was not done for the second round of interviews, with the result that the figures given for these are undoubtedly too high. The more detailed analysis provided in subsequent tables and the comments that follow here are all based on the first round of interviews, and this should be kept in mind. Time prevented any close scrutiny of the data from the second round of interviews: in any case, it is quite unlikely that the results reported here would be significantly different had this data been included in the analysis, given the very considerable number of obligatory contexts for article usage in a single interview.

4.13.2 Frequency of Realization

The first point that has to be made in relation to both the definite and indefinite articles concerns their low rate of realization. So far, it has not been possible to identify by hand in the transcripts the obligatory or optional contexts for their production. On the evidence of Table 4.13.1 alone this looked rather low. However, in order to provide some additional approximate idea of how frequently or infrequently articles were produced, the following calculations were made. The number of nouns occurring in a given interview was tallied from the lexicon for that interview. Then the number of articles produced was expressed as a percentage of this figure. Obviously, since not all nouns would be obligatorily preceded by non-null articles (they could be legitimately preceded by null determiners, or by adjectives or quantifiers) this procedure can only provide a rough idea of the suppliance of articles in obligatory contexts. However, even assuming that half the nouns produced should have been marked with an article, the realization rate so calculated is quite low for most of the informants.

II	I
+SR +HK	-SR +HK
IV	**III**
+SR -HK	-SR -HK

This generally low rate is reduced even further in the case of an, with a total of a mere seven tokens in the whole cross-sectional corpus.

A closer examination of these tokens makes it clear that the phonological rule for an is just beginning to establish itself in these speakers. The one token for Ludwiga J. occurs in the phrase half an hour. One of the two tokens produced by Andrzej J. shows evidence of hypercorrection:

> they... will get... er... an... an good work in... their job
>
> - aj.1: [708] -

While in the case of Phuc, the one Vietnamese speaker to produce tokens of this form, three out of four come in the phrase an interpreter.

A further point of interest is that all valid tokens of an in the first round of interviews occur in front of words beginning with */i/*. This suggests that the most favourable phonological contexts for the operation of the *n*-insertion rule for the production of an are high front vowels. From an articulatory viewpoint this makes sense, since an unmediated transition from a mid or—as is more likely in learner language—a low back vowel to a high front one would be the most difficult transition to manage.

4.13.3 Syntactic and Semantic Environments

The suppliance of articles in the present corpus seems to be conditioned fairly strongly by semantic, and to a lesser extent, syntactic factors.

4.13.4 Semantic Environments—The Definite Article

In the case of the definite article an analysis similar to that conducted by Huebner was carried out [85]. Huebner, following a model proposed by Bickerton, argues that noun phrase reference can be accounted for in terms of two binary features *+/-Specific Referent* $(+/-SR)$ and *+/-Assumed Hearer's Knowledge* $(+/-HK)$. This provides the following semantic field:

In English, noun phrases in Sector I are marked with the, while those in Sector IV—that is, generics—can be marked with the, a or pluralized and

marked with a null marker. Noun phrases in Sectors V and VI are marked with a or the null marker.

The analysis presented in Table 4.60 defines a fifth category—that of proper names or formulas. Semantically, these are a subset of the cases covered in sector II. However, given the status of formulas or monomorphemic chunks in the learning process it seemed advisable to allot them a separate classification: there is no guarantee that the article in such utterances is analyzed as such and is therefore a meaningful unit. (Huebner's solution to this problem was to exclude what he judged to be proper names and idioms from his analysis. This may be permissible if the focus of the study is primarily semantic, but from an overall developmental point of view it is clearly unsatisfactory, since formulaic language is arguably the seedbed of propositional language).

Phrases such as the city, the Communists and the hostel were classified as proper names, and utterances like the best country, the next day and the same were classified as formulas. As always there were borderline cases that had to be decided largely on intuition. A more general problem with making judgements of semantic intent on discourse based evidence is that lack of cohesion in the discourse itself can be a complicating factor. Models like the one above may appear to be applicable to all varieties of language, but in practice are a good deal easier to apply to native speech, with its well articulated discourse features, than to learner language.

As is clear from Table 4.60 the majority of the definite articles produced are as markers for proper nouns or generics. In practice, it can be quite difficult to distinguish between these two categories if the generics are, as is frequently the case, non-idiomatic. Some examples of borderline generics would be the following:

> the food here is not expensive
>
> - vinh.1: [273] -

> I... I have... I've learn... I have, um, read [/rid/] the book and, er,... er, take on paper, er, w... words and I learn words,
>
> - mm.1: [58–9] -

Given that generics do not involve specific reference, proper nouns involve it by default rather than actively (that is, the referent is effectively a member of a unique set), and the function of the definite article in a formula is empty or redundant given the semantic content of the formula (e.g. the best), this means that a preponderance of the definite articles produced by learners up to quite a high level of proficiency do not actually involve positive specific reference.

This pattern of usage is quite significant, since it means that in effect the definite article is in most cases in the present corpus not really used as a discourse marker.

When the definite article is used to indicate positive specific reference, there is evidence of confusion over the feature of *Assumed Hearer's Knowledge*. Thus, we find a number of cases where Sector III noun phrases would be acceptable to native speakers. Examples are:

> mm... firs(t) time... um... have er sik boas, er, in the, er, raiver ...the raiver... er
>
> - sang.1: [703–4] -

> In one week... er... the material... came
>
> - ka.1: [281] -

(It should be noted that the marked noun phrases in these cases have no prior referent—their possible acceptability depends on them being interpreted as idioms). In addition to these, however, there are some clear cases of unacceptable tokens, such as:

> No water... mm... every very tire(d), um... but, er... the s(h)i(p) of Thailan(d), er, ro(f)... after ro(f), er... they ta(ke) some water and ri(ce)
>
> -sang.1: [764–6] -

> I don't know yet... /... if, er... er, it will be possible, er, to work... and er..., er... not have [?] the, er, special er... electronics course
>
> - aj.1: [675–9] -

Amongst the Sector III tokens there are also some interesting cases where the speaker's confusion about which surface realization to assign to noun phrases in this category is quite patent:

> How long, er, how long, er, you work a job, er... this job
>
> - minh.1: [469] -

> people they, er, sell the something around... around the camp
>
> - vinh.1: [272–3] -

it was not possible... the one person, er, must stay in Poland

- aj.1: [590] -

and I was working in the, er, this brown coal mine

- mm.1: [189] -

Yes, it's, er, social problem because in the... a few last years

- mm.1: [479–80] -

In the mastery of the semantics of the definite article there is evident difficulty in simultaneously controlling the features of *Specific Reference* and *Assumed Hearer's Knowledge*, once the feature of *Specific Reference* becomes a factor in production of the definite article—that is, once the speaker's noun phrase marking extends beyond names, formulas and generics.

The Question of Sequencing

It is not really clear whether we would be justified in speaking of a semantic developmental sequence for the definite article.

In Huebner's analysis, which is based on the semantic quadrant presented above, his informant's progress seems to be characterized first by a series of inappropriate hypotheses—one founded on the notion of topic marking, another on global marking of noun phrases—and then by a gradual process of elimination of marking in those environments which are ungrammatical in English. This process is characterized by stops, starts and backsliding, and seems on the surface very plausible.

Such an interpretation is based, however, on the assumption that the features used in the analysis exist independently of the developing language which embodies them. That is, Huebner assumes that the learner is merely applying a set of pre-existing universal discourse features in different ways until he gets the right fit rather than actually discovering the semantic properties of the article system in English as he goes along. This effectively disallows any interplay between form and function: functions are already defined and it is merely a question of getting the forms to index them in an appropriate way.

There are problems with such an approach. Firstly, the features involved are both complicated and rather abstract, and it is difficult to envisage a psychologically real process of hypothesis testing involving them. Second, it is difficult to reconcile a developed set of oppositions of the type presented in the semantic quadrant with formulaic—semantically opaque—language. With the relevant semantic categories already established, such language is

anomalous. This leads to Huebner excluding from consideration such things as idioms, proper nouns and commonly used expressions. Yet elsewhere Huebner himself argues against the exclusion of such language forms from any analysis of developmental processes, and we have abundant evidence ourselves that formulaic material is instrumental in the evolution of propositional language [139].

An alternative approach to Huebner's which would admit all the data to hand is to characterize the learner's task in a different way. Rather than comprising an attempt to fit hypotheses to a pre-existing framework the learning task can be viewed as a process whereby this framework is actually developed. Such an interpretation is much more in line with other developmental processes we have examined—for instance, the growth of the semantic tree which characterizes the pronominal system. In addition, it allows for interplay between form and function.

In this approach, the definite article would enter the system in proper names and other forms of formulaic language where specific reference is in fact a semantically empty notion, given the one member sets involved. The feature of *Assumed Hearer's Knowledge* would come to be associated with the definite article through generics. (These are considerably more numerous in the present corpus than in the data from Huebner's informant).

As the learner's discourse became more particularized and noun phrase sets became contrastive—that is, multi-membered—the feature of *Specific Reference* would emerge. The problem for the learner would then be to decide which combinations of *Assumed Hearer's Knowledge* and *Specific Reference* were valid correlatives of the definite article in English. This is more or less the process we have commented on, and, as Table 4.60 shows such an interpretation is consonant with the data in the present study. The amount of non-formulaic data that we have on which to base any hypotheses we may wish to form is not great, and it would be rash to be over-assertive about the above conclusions. But regardless of their status, given the preponderance of proper names and formulas in the present corpus, we simply cannot afford ourselves the luxury of ignoring formulaic language, as Huebner does: too much of the data is in precisely this form.

Evidently, the process described above, or indeed the process described by Huebner is not an inevitable one. In his longitudinal study, Huebner describes a phase in which his informant simply marks all noun phrases with the definite article [85]. In the present study, this phenomenon is to be observed in two of the Vietnamese informants, Canh and Long. These informants supplied a higher percentage of articles before nouns than any other informants in the study (see Table 4.52), but analysis of their output revealed that their marking of noun phrases with the definite article was quite indiscriminate. While some sectors of the quadrant are more heavily represented than others, numerous examples of marked noun phrases in every sector are in evidence:

Sector I:

I think there is ju(s)t, er, the place they want to test weapon(s)

- canh.1: [736–7]

the father must eat before the children, yeah

- long.1: [963–4] -

SectorII

I can understan(d) the life Australian living

- canh.1: [584] -

he help take the hand-bag he come see me an(d) give back the ticket for the bus

- long.1: [170–1] -

Sector III:

If I, ah... I ha... I have the girlfriend... I want to see her, oh, ve(ry) difficul(t)

- canh.1: [1339] -

to the city an(d)... I go out to the, er... bus station, you know, I see my friend on the bus

- long.1: [167–8] -

Sector IV:

they haven't got the government, they just got a Junta

- canh.1: [781] -

they get we somewhere too... because they have the trouble before in Vietnam

- long.1: [247–8] -

Whether in their case it is a phase or the end of the road as far as developments in the system of articles are concerned is a moot point. Given the fact that both informants had been in Australia for three to four years at the time of the interviews and were amongst the more communicatively effective informants in the sample, the latter possibility—that of rule fossilization—seems the more likely of the two.

4.13.5 Developmental and Variational Features

We have discussed two possible views of the development of a rule system for target-like production of the definite article in English—a hypothesis fitting model and a developmental one. Both approaches suggest that there are regularities in the process of acquisition of the definite article. Regularities in an acquisitional process, however, do not necessarily indicate that the feature being acquired is a developmental one. We have seen that there is evidence for the existence of ordered environments for the suppliance of the copula, yet this is almost certainly a variational feature [128]. Cases like those of Canh and Long emphasize the next question to which we have to turn—that of whether the definite article is an acquisitional or variational feature.

Morpheme order studies assume that articles, like all other morphemes are developmental features. In doing so they have given no consideration to the underlying semantics of article use except in so far as this appears to coincide with target language patterns. As Huebner points out, target-like usage patterns may not actually be the product of target-like semantics [85]. The present data supports Huebner's contention. Speakers like Canh and Long have the highest rate of suppliance of definite articles of all the informants in the sample, and would therefore most closely approach morpheme order criterion for acquisition of this item. On the other hand, their underlying semantic systems appear to be among the least developed of all for those informants who use definite articles. For this reason, amongst others, we cannot adduce the evidence presented in morpheme order studies as evidence in favour of the classification of articles as developmental features.

Ultimately, the question of whether the definite article is a developmental or variational feature cannot be definitively settled until independently motivated developmental sequences have been established for English. If we make some rough assumptions about what these might be, however, the status of the definite article as a developmental feature becomes quite doubtful. Thus, in negation Canh appears to have a reached a fairly advanced stage of development, employing only standard preverbal negators in a wide range of contexts and even suppleting some into any in the scope of negation. In terms of the verbal system, he uses a wide range of irregular pasts, produces more regular pasts than other Vietnamese speakers, marks some verbs with third singular *-s*, and produces some gerunds. If the definite article were a developmental feature we could reasonably expect him to be at a more advanced stage than the global noun phrase marking stage he in fact appears to be at. For Long, the discrepancy between his pattern of usage of definite articles and other aspects of his grammar is not so great. Nevertheless, if we are talking about a developmental feature one dubious case is enough. In addition to these considerations, the most detailed description that we have of the acquisition of the definite article over time—namely Huebner's—

suggests that the acquisition process is by no means linear, or constrained in the way that one would expect the acquisition process for a developmental feature to be.

The case of the definite article *vis-a-vis* the developmental/variational distinction is an interesting one. Some of the conditions for a developmental feature are met: namely that a postulated developmental sequence can be independently motivated in that the development of a workable set of semantic distinctions must take place, in a manner analogous to the development of processing prerequisites necessary for the acquisition of word order rules. And then there are the implicational type patterns of Table 4.60. On the other hand, there are the anomalies just mentioned. Clearly, this area is one in need of further research.

The Question of Transfer

While neither Polish nor Vietnamese has an indefinite article, Vietnamese uses a postposed demonstrative adjective to indicate prior reference. An example would be:

Anh có anh không?—"You have brother not-Q?"

Anh âý lãm gĩ?—"Brother that work what?"

In Chapter Two a hypothesis was advanced that interference on the morphological level was likely to be higher than say, on the clausal level. Given the existence of a definiteness marker in Vietnamese we might therefore expect that this would predispose Vietnamese speakers to be more sensitive to the possibility of such a marker in English. Tables 4.13.1 and 4.56 show that the absolute rates of suppliance and the percentage of nouns marked with the definite article is, in general, slightly higher for the Vietnamese speakers than for the Polish speakers. It is hard to tell if this difference is really significant, much less attribute it entirely to transfer. Nevertheless, the possibility cannot be discounted. Elimination of proper nouns and formulas from consideration accentuates the higher rates of suppliance for the Vietnamese speakers and therefore enhances this possibility. Nevertheless, it should be borne in mind that the differences in suppliance of the definite article are nowhere near as dramatic as those, say, for copula insertion. If there are first language transfer effects they would seem to be conditioned by functional considerations: whereas the copula is semantically vacuous and can be safely ignored, some means of indicating definiteness will have to developed at some point for communication to be effective.

In regard to the indefinite article, the low rate of suppliance for both groups may be due, amongst other factors, to its absence in the first language. In this case, since both languages are the same in this regard, a comparison is not possible: it would be interesting to see how a comparable

group of speakers of a language which does have the indefinite article fared. (See below for further discussion in relation to data from Spanish speakers).

Syntactic Environments—The Indefinite Article

As Tables 4.13.1 and 4.56 show, for most informants in both gross and percentage terms the definite article is the more commonly produced. There are probably several reasons for this. It is phonetically more salient than the indefinite article; it occurs in many frequently used idioms and proper names and can therefore enter the system in a wide variety of formulaic contexts, and in semantic terms it involves more plusses than minuses—that is, it is less vacuous than its indefinite counterpart. We have seen one consequence of the salience of the definite article above, in its choice as a global noun phrase marker by Canh and Long. If a determiner is to be overgeneralized or used indiscriminately as a nominal marker it will almost certainly be the definite article: in the present corpus there is virtually no evidence of the indefinite article being used in this way. While the acquisition of the definite article involves a process of overapplication and subsequent constraint, the acquisition of the indefinite article seems to involve a more straightforward process of emergence in which it supplants either null markers or the definite article. For these reasons a semantically based analysis of the acquisition of the indefinite article would not be very enlightening.

The procedure that has been followed with the indefinite article has been rather to look at its emergence in terms of syntactic environments. This is a process that has been hinted at in the discussion of the definite article, where it was suggested that it enters the system in formulaic language.

The analysis of the acquisition of the indefinite article is based on hypotheses put forward by Pica. In a study of eighteen Spanish speakers acquiring English under different conditions Pica proposed that the indefinite article "followed a developmental sequence characterized by its accurate suppliance in the following implicationally related environments:

1. Isolated units such as a few, a little, a lot, half an hour, once a month, twice a week.

2. The same isolated units embedded in predicate structures, e.g. have a little (NP), ate a lot (of NP), visits once a month.

3. Noun complements and direct objects in predicate structures such as is a friend, have a son, read a book.

4. Noun objects of prepositions such as in a minute, with an accent [139].

Table 4.62 presents the results of Pica's analytical framework applied to the present corpus. The results obtained are somewhat less regular than Pica's, although the picture would improve somewhat if categories (1) and

(2) were collapsed (as Pica notes, differential rates of suppliance for these two categories can be explained in terms of increasing structural complexification of the speakers' language—which is not a factor of very direct relevance to the production of articles *per se*) [139].

It is difficult to determine why the present results do not replicate Pica's. One possibility is that informants in her study were developmentally more advanced than those in the present study. Another possibility has to do with the language backgrounds of the informants in the two studies. Spanish, unlike Polish or Vietnamese, has an indefinite article. The semantics of the Spanish indefinite article are similar, though by no means identical, to those of the English one (English requires suppliance in contexts where Spanish may not [173]) and this may have led to a higher rate of suppliance in Pica's data. That first language morphological patterns can manifest themselves in a second language is a proposition which has been advanced at several points throughout this study: specific evidence for interference with articles was found by the author in the learner language of Turkish speakers, where the form one (the Turkish number and indefinite article are identical) was preferred to a [92]. (There is, of course, also the evidence relating to the definite article presented above).

Whatever the reasons for the lack of detailed concordance in the two sets of findings, the principal finding—that forms enter the system through formulas—is the same for both studies and adds to the already considerable body of evidence that formulaic material is the seedbed for propositional language: in the present study only two out of twenty informants produce articles in environments (3) or (4) without showing any evidence of isolated units.

One interesting sidelight on this principle is that through a categorial change certain isolated units can actually result in the spread of articles to environments (3) and (4). This is not commented on by Pica, so it is worth outlining.

The isolated unit concerned is a little. The phenomenon occurs in two informants—Ludwiga J. and Mieczyslaw M. In the case of the first informant, Ludwiga J., a little is used as an adverb (Pica's environment 2):

I know this problem a little too I

- lj.1: [239] -

As Table 4.62 shows, such category (2) usages constitute the majority of her article tokens. However, through the use of a little as an adjective Ludwiga J. actually extends her production of article tokens into categories (3) And (4):

I'm no a little leader

- lj.1: [519–20] -

and this standing one woman and have a little piece of bread... we buy and all shop was empty

- lj.1: [339–40] -

he sitting all day in bed... can't work and have poor family in, ah, a little country

- lj.1: [256–7] -

It is difficult to tell whether a little is still monomorphemic in these examples. In any case, this is not really crucial, since one would presume that the categorial and environmental extensions effected would lead sooner or later to a reanalysis of the unit into its constituents of article plus adjective/noun.

That this is not merely fortuitous and idiosyncratic process can be shown from the data provided by Mieczyslaw M. This informant uses a little as an isolated unit:

but, er, a bit, er, ah, um, a litt(le) more in Zgorzelec

- mm.1: [288–9] -

and as an isolated unit in a predicate structure:

and, er, and, er... in television, er, last time in Poland was a litt(le) about Australia

- mm.1: [146–7] -

The process of transition is very clear in Mieczyslaw M.'s case. The recategorization of a little from adverb to adjective may be assisted by the optional movement of the adverb to a position between verb and object. Thus, we have sentences like:

Eh, yes I think, er, I speak German a litt(le), er, a little better as English

- mm.1: [862–3] -

and:

and, er, he speak, er, a little English but not correct

- mm.1: [100] -

Placement of the adverb in this position renders it analyzable as an adjective, as in the token above, or the following one:

and many people speak a litt(le) Russian

- mm.1: [877] -

Then there are several definitely adjectival usages of a little, such as:

Yes, er, in the, er, this... er it was, eh, a lit(tle) town, er, about, er, ten thousand inhibitants

- mm.1: [220–1] -

Yes, and near, or, the lake was, er, a litt(le), um... a litt(le) houses

- mm.1: [298–9] -

These occur in category (4) environments as well:

long time ago but, er, I lived, er... in, er, a little town

- mm.1: [75] -

In Mieczyslaw M.'s case, we have evidence that adjectival a little is definitely analyzed. This is rather neatly illustrated by the following example, in which big is the adjective substituted:

don't know... is wat... not lake, only a litt(le), er ... like, it's, er... a big, er, water

- mm.1: [290–1] -

The whole of the above process is a model example of how a monomorphemic utterance is broken down through a process of repeated use in different contexts, and of the accompanying categorial adjustments that may occur during such a process.

Within the hierarchy of environments defined by Pica there appear to be further hierarchies. Thus, for category (3) environments there are certain verbs which favour the appearance of the indefinite article. Tables 4.63 and 4.64 show these—have, get and be constitute the most congenial environments for the production of the article.

Informant	is	zj	ks	jb	jr	ka	bb	es	aj	kb	lj	mm
the		5	2	6	55	22	6	17	24	34	17	35
a				4	10	3	4	6	30	7	40	32
an									2*		1	

Table 4.52: Distribution of Articles: Polish Informants—Interview 1

4.13.6 Semantics of the Indefinite Article

As mentioned in 4.13.3, the indefinite article is not subject to the semantic vagaries of its definite counterpart—it is not the chosen form for non-standard marking of noun phrases, and its emergence is slow but seems to be relatively orderly. All this could be seen as a consequence of the heavy contextual dependence—in both syntactic and verbal environments—which has been outlined above. With Phuc, the speaker whose use of the indefinite article is probably the most developed of the informants in this study, there are in fact some indications of non-standard semantics for the item. In Phuc's output there is some indication that the indefinite article is not consistently interpreted as an analogue of one. This is evident in the following examples:

he doesn'(t) make a...friends with Vietnamese

- phuc.1: [48] -

Yeah...but he tol(d) me tha(t)...I should have a one job

- phuc.1: [79] -

4.13.7 Summary

To recap, for most informants the rate of article suppliance appears to be quite low. Both the definite and indefinite articles enter the learner's system through proper names, idioms and other formulaic material. In the acquisitional process the definite article is the principal nominal marker and its acquisition is characterized by a process of semantic discrimination and hypothesis fitting that can proceed at very different rates for different learners. The acquisition of the indefinite article seems to be a somewhat more stolid and orderly process, in which null marking or non-standard usages of the definite article are gradually supplanted. There is some evidence that transfer phenomena may affect this area of learner grammar, but further research is required before more definite conclusions can be arrived at.

Informant	is	ks	zj	jr	ka	jb	bb	es	aj	kb	lj	mm
the	9	21	10	43	19	11	39	20	40	29	21	54
a	8		2	7	7	12	3	5	48	6	46	10
an									2			

Table 4.53: Distribution of Articles: Polish Informants—Interview 2

Informant	Van	My	Duc	Dung	Minh	Hoa	Sang	Vinh	Tam	Canh	Long	Phuc
the			7	15	4	11	20	33	36	258	261	116
a		1	16	13	5	4	21	14	14	39	40	54
an												2

Table 4.54: Distribution of Articles: Vietnamese Informants—Interview 1

Informant	Van	My	Duc	Dung	Minh	Hoa	Sang	Vinh	Tam	Canh	Long	Phuc
the	8	16	27	14	28	11	6	36	50	56	25	49
a	12	23	14	6	6	3	18	26	16	27	4	29

Table 4.55: Distribution of Articles: Vietnamese Informants—Interview 2

Informant	is	zj	ks	jb	jr	ka	bb	es	aj	kb	lj	mm
the		2.1	0.8	2.9	11.7	7.7	1.2	3.7	5.6	8.4	2.0	6.0
a(n)				1.9	2.1	1.1	0.8	1.3	7.1	1.7	4.7	5.5

Table 4.56: Percentage of Articles Compared with Nouns: Polish Speakers—1

Informant	is	ks	zj	jr	ka	jb	bb	es	aj	kb	lj	mm
the	5.6	7.7	4.0	14.7	6.9	4.6	9.4	7.7	10.3	8.2	3.2	15.0
a(n)	5.0		0.8	2.4	2.6	5.0	0.7	1.9	12.9	1.7	8.1	2.8

Table 4.57: Percentage of Articles Compared with Nouns: Polish Speakers—2

Informant	Van	My	Duc	Dung	Minh	Hoa	Sang	Vinh	Tam	Canh	Long	Phuc
the			2.1	0.4	2.6	3.8	6.7	11.6	8.1	31.6	20.7	17.8
a(n)		1.4	4.8	4.6	3.2	1.4	7.0	4.9	3.2	4.8	3.2	8.6

Table 4.58: Articles Compared with Nouns: Vietnamese Speakers—1

Informant	Van	My	Duc	Dung	Minh	Hoa	Sang	Vinh	Tam	Canh	Long	Phuc
the	10.7	4.3	10.6	6.3	17.6	4.9	2.5	17.1	15.4	20.9	18.8	20.8
a(n)	16.0	6.2	5.5	2.9	3.8	1.3	7.5	12.4	4.9	10.1	3.0	12.3

Table 4.59: Articles Compared with Nouns: Vietnamese Speakers—2

Informant	is	zj	ks	jb	jr	ka	bb	es	aj	kb	lj	mm
Proper N. Formula		5	2	6	48	13	4	12	21	26	10	25
Generic -SR/+HK					2	4	2	2	1	3	2	
+SR/+HK Oblig.					3?	5		3		2	4	4
+SR/-HK Var.						5			4	2	1	6
-SR/-HK Non-Std												

Table 4.60: Functions of the Definite Article: Polish Informants—1

Informant	Van	My	Duc	Dung	Minh	Hoa	Sang	Vinh	Tam	Canh	Long	Phuc
Proper N. Formula			4	15	3	3	15	17	26	*	*	85
Generic -SR/+HK			3		1	8	3	8	3	*	*	5
+SR/-HK Var							3	3	4	*	*	4
+SR/+HK Oblig								5	3	*	*	20
-SR/-HK Non-Std										*	*	

Table 4.61: Functions of the Definite Article: Vietnamese Informants—1

Informant	is	zj	ks	jb	jr	ka	bb	es	aj	kb	lj	mm
be				1			3		10	1	3	5
have					2				9	1		
do					1							
get									2			
know									1			
give									1			
speak												2
0									1	2		

Table 4.62: Environments for "a/an" in Noun Complements: Poles

Informant	is	zj	ks	jb	jr	ka	bb	es	aj	kb	lj	mm
Isolated Unit				2	1	3	1		2	3		6
I.U. in Predicate				1	3				2	1	36	16
Noun Comp/ Direct Obj.			1	4			3	4	26	2	3*	7*
Object of Preposition					2			2		1*	1*	3

Table 4.63: Environments for the Definite Article

Informant	Van	My	Duc	Dung	Minh	Hoa	Sang	Vinh	Tam	Canh	Long	Phuc
Isolated Unit	1		3		2		3	3	4	7	4	1
I.U. in Predicate			2					2	2	10	6	6
Noun Comp/Direct Obj.			13	8	4	2	14	8	6	20	30	38
Object of Preposition			1	2	17			1	2	2		13

Table 4.64: Environments for the Definite Article

Informant	Van	My	Duc	Dung	Minh	Hoa	Sang	Vinh	Tam	Canh	Long	Phuc
have			4		2		6	4	1	3	10	5
get				5		1		1	1	8		6
be				1		1		1		2	1	10
send							2		1		2	1
look for			3						1	1		
give				1							2	4
take			1								1	
buy					1		1					
work					1	2						
make											1	4
OTHER			2				1	2	1	5	2	7
0			3		1	2			1	2	2	1

Table 4.65: Environments for "a/an" in Noun Complements: Vietnamese

NOTES:

1. In the asterisked cases marking of NPs was indiscriminate and voluminous and occurred in tokens in all categories.

4.14 Quantifiers

4.14.1 Distribution of Quantifiers

Tables 4.66 to 4.14.5 provide counts for the distribution of quantifiers within the two interview groups, provide counts for the occurrence of quantifiers according to *ASLPR* level and language background. Though not quantifiers themselves, words such as everything, which consist of a quantifier bound to an noun phrase, have been included, since in some cases a quantifier may first appear in this kind of bound environment.

As Tables 4.66 to 4.14.5 show, on a distributional basis the first eight ranks of the tables are occupied by the same set of quantifiers, with the exception of Table 4.67, where few displaces lot.

Moreover, the rank orders in all cases are very similar, with the only really noteworthy difference being the higher position of some in the tables based on data from the Vietnamese speakers.

As regards total numbers of occurrences of each item within each interview group, there is also considerable regularity. Thus, for the Polish speakers, the first eight most frequent items are the same for both the first and second interviews; for the first five items the order of frequency is identical. They are many, all, much, some and every, in that order—the same five most frequent items on a per informant basis. For the Vietnamese speakers, the gross counts are not quite so tidy but nevertheless of the first six most frequent items, five are the same. The five are all, every, many, much, and some, once again.

4.14.2 Developmental Trends

At first sight, with the five main quantifiers (and their bound variants), so widely distributed, and most of the others so narrowly distributed and infrequent, it appears that there may not be a great deal more to say about them, certainly in regard to any possible orders of acquisition that might be deduced from the cross-sectional data. There are, however, a number of points raised by a more careful scrutiny of the data, and several indications of possible developmental trends. The following discussion is based largely on careful scrutiny of the data from the first round of interviews. The points that will be made seem to be largely supported by the tabular data from the second round of interviews (that is, Tables 4.67 and 4.14.5).

Distributive and Global Quantifiers

First, it is worth noting that the most widely distributed quantifier is no, or some bound variant thereof. No is what might be termed an absolute quantifier. It expresses an unequivocal proposition. When non-count noun phrases are quantified by no, this proposition applies to them globally. When no quantifies count noun phrases, as in the bound form nobody, it is distributive, in that it applies to every possible member of the quantified set. In this regard, it is similar to all and every, and distinct from quantifiers like some and many, which involve degrees of applicability. Learners with a limited set of quantifiers almost always have as members of this set representatives of what could be described as the semantic poles of all and none, as can be seen from the tables. (See below for comments on early tokens of much and many). In the present data, the negative pole is realized as no, or one of its bound variants, and the positive pole as all or every, or one of its bound variants. Acquisition of these polar quantifiers, then, appears to be a precondition for the acquisition of quantifiers of degree, like some, much, and many.

The Case of 'Every'

As noted, the semantic complement of no is either every or all. One apparent anomaly in the data is the fact that both these two fairly close semantic relatives are so widely distributed and frequent. This kind of duplication runs contrary to the form-function principle. An examination of the examples in which every occurs, however, reveals that it overwhelmingly figures in adverbial phrases of time—every day, every week, and so forth. Amongst the more advanced learners there are some cases of every appearing in other contexts, and one informant, Ludwiga J., does not use it at all in adverbial phrases of time; but these are certainly the exceptions. Thus, there may be a certain degree of formularity in the use of every, even at a quite advanced stage. If this is so it might help to explain another phenomenon, namely that unlike

many other functional items every does not occur more frequently in the output of more advanced or proficient learners. The counts for every remain relatively stable throughout the sample. This is what one might expect for a formulaic item whose production continues to be stimulated by specific triggers rather than governed by principles with a gradually widening ambit. Apropos of this last observation, it is worth noting that all is produced, in most cases, more frequently by the upper-end informants.

The case of every is instructive for another reason. In the first instance, the environment for every, categorially a quantifier, is an adverbial phrase. As evidence to be adduced in regard to other quantifiers, such as some and much will demonstrate, where it is possible, quantifiers tend to emerge from a previously acquired adverbial matrix. This constitutes an important acquisitional principle.

4.14.3 'Statistical' Quantifiers

'Very'

It was mentioned in the previous section that there is evidence for the proposition that quantifiers emerge from an already acquired adverbial element. This is particularly apparent in the case of the first of the "statistical" quantifiers (that is, quantifiers which do not apply to each and every member of the quantified set). The first such quantifier is a non-standard form in its capacity as a quantifier—namely very. (The assertion that very is the first statistical quantifier to be used is based on the evidence of Irena S., who does not produce any of the other forms in this class). Examples of very as a quantifying adverb or quantifier are the following:

Mus(t) very learn

- is.2: [531] -

He very, er, help me

- jr.1: [605] -

I'm very li(ke) my country

- hoa.1: [643] -

I very li(ke)... I like very much, er, Australia

- minh.1: [324–5] -

Tokens of very as an apparently adjectival element can also be found:

car is not problem... is... have not money... is very problem

- bb.1: [779] -

It could be argued that the first set of examples are really no more than adverbs with a quantifying function. However, the early categorial overgeneralization of very does also lead to explicit occurrences of the word as a quantifier:

is very money

- bb.1: [779] -

is very... very much engineer(s)

- bb.1: [1009–10] -

In a previous study, the use of very as a quantifier was also observed [92]. In the last example from Barbara B., the self-correction provides a rather neat vignette of a probable developmental trajectory, in which very is refined into very much, or too much, which in turn, it would seem plausible, is later reduced to much alone (see below for further discussion). Some support for this argument (as well as an example of the persistence of very as a quantifier) can be found in the fact that in **lj.1** very is used as a quantifier, as in the following examples:

I very love children

- lj.1: [589] -

it was very, very supermarket(s) in Warsaw

- lj.1: [333] -

while much does not appear at all: in other words, there is a complementary relationship between the two forms. Despite the example of Ludwiga J., we could generally expect tokens of very as a quantifier to have largely disappeared by about level 1 on the *ASLPR*.

'Many'

Of the other quantifiers with wide distribution, many exhibits an increase in frequency of occurrence towards the upper end of the range. This is presumably explicable simply in terms of the increased fluency or volubility of those speakers higher on the *ASLPR*. A further point worth noting is that in the case of the lower-level Vietnamese informants, Duc, Dung and Minh, where it seems to have emerged prematurely, many occurs not as a quantifier but as the WH-word, how many, in two out of five tokens for Dung, and all tokens for the others. It is only in Hoa, and the informants ranked after her, that many appears consistently as a quantifier. The early appearance of many in a WH-word perhaps provides an indication of one way many might enter a learner's system.

Two of the Vietnamese speakers, Tam and Long, realize many as too many, in all, and nearly all, cases respectively. This lack of a distinction between the quantifier in its base form and in comparison can be more generally observed in examples with much. A possible explanation for the binding of too and much will be offered further on.

'Some'

Another quantifier which seems to exhibit a developmental pattern is some.

Amongst the Polish speakers, it is virtually absent until level 1+ on the *ASLPR* is reached. (The fact that Mieczyslaw M., the most highly rated Polish speaker produces only one token of some in two interviews once again raises the question of how well developmental features and oral proficiency correlate, and should be kept in mind).

Some appears principally as an indefinite quantifier, or plural indefinite article. On occasion, it is used with singular count nouns as well:

> we readet, er, about Australia, er, some books
>
> - aj.1: [269] -

> er... now wan... when we woos [will?] like, er... er, to... to speak about, er... some, er, very big problem
>
> - aj.1: [475–6] -

> you... you know and, er, we have only some book(s)
>
> - lj.1: [474] -

> they need some money
>
> - vinh.1: [567] -

ah, they have, ah... they make some problem(s)

- canh.1: [1866] -

Less frequently, it is used in its other sense, with a specific although unspecified referent (in Huebner's terms, *+Specific Reference*, *-Assumed Hearer's Knowledge* [85]):

Ah, er... er, some of them, er... er, got job(s)

- aj.1: [713] -

because some people, you know, they don't have education

- vinh.1: [400] -

Where only one of these senses is present in the output of a particular informant it is the former. This is probably a reflection of the developmental sequence in regard to these two senses of some. It is also worth noting that in all but one case speakers who produce some also produce tokens of sometimes, which may very possibly mean that the form sometimes provides the developmental matrix for some. Orthography aside, this is precisely what was observed for every, where the quantifier really existed as a bound element in an adverb, and had established itself outside of this environment in only a few cases. (Although the parallel is not exact, it would be interesting to look at all in relation to always). The final point that should be made about some is that it is both more widespread and more frequently produced amongst the Vietnamese speakers. The reasons for this are not clear: either first language patterns or differential patterns of exposure for the two groups may be an influence here.

'Any'

Any receives some discussion in 4.10 so it will only be dealt with briefly here. Amongst the Polish informants, in the first interview at least, there are no certain tokens of any. As pointed out above, this is possibly a product of language-specific features in the Polish determiner system. Likewise, bound forms of any, such as anything or anywhere are not produced either. Amongst the Vietnamese speakers, any is somewhat more in evidence, especially in those informants who rated over 1+ on the *ASLPR*. This fairly neatly reflects the pattern of occurrence of some, to which any is related. As a form any has two basic functions. In the scope of negation, and in interrogation, it is the suppleted form of some—that is, an indefinite quantifier or determiner. Its other, not unrelated function, is as a referent for any single possible member of a given set ("Think of a number, any number"). Speakers who

use any produce tokens with both functions, in roughly equal numbers. On the analogy of the pattern of acquisition for some, it would be reasonable to expect any to appear first in one of its bound forms, such as anything. However, there is no particular support for this in the data.

'Much'

While much is not an item which exhibits such an evident developmental pattern as some, the pattern whereby a quantifier appears first in an adverb or adverbial phrase, and then emerges from this context occurs at least in a subset of the data—in the output of the Vietnamese speakers. Thus, the first five Vietnamese speakers who use much use it either in an adverbial phrase, in such expressions as the following:

> Drink beer, yeah, drink beer... too much... now no
>
> - dung.1: [127] -

or as a bound element in the WH-word how much (see Table 4.67). Speakers further up the *ASLPR* in the data examined (that is, Tam, Canh, Long, and Phuc) may produce tokens of adverbial much, but they also use much, or some variant, such as too much, as a quantifier.

> Er, yes, ah... since I may helf my friends, you know ... who can'(t) speak much English
>
> - phuc.1: [607] -

The appearance of much as a quantifier, while it may mark a developmental step forward is also the occasion for further possible non-standard features to emerge in the learner's speech. This is because of the idiomatic way much is used by native speakers. Thus, in Standard English much is generally only used, in casual speech anyway, in the scope of negation, or in interrogation, or if preceded by an adverb—that is—while we say I don't have much money or He has too much ambition, or Is there much bread left?, we do not normally say things like They had much fun. Avoidance of the latter structure is normally achieved by using an alternative lexical form such as a lot of, or a great deal of. This idiosyncratic constraint on much has to be learnt by anyone who begins to use the word as a quantifier, and mastery of it could therefore be said to constitute a further developmental stage in the use of much.

A number of informants do produce tokens of a lot (of), which is a precondition to acquisition of the constraint. These are Jan R., Ewa S., Krystyna A., Ludwiga J., and Mieczyslaw M., amongst the Poles, and Tam, Canh, Long and Phuc, amongst the Vietnamese. Of these, Mieczyslaw M.,

Canh, Long and Phuc provide evidence that they are quite sensitive to this constraint. Thus Mieczyslaw M. only uses much post-adverbially or in the scope of negation:

> er, it...it was difficult for me because, er, um, not much, um, people, er, know this language
>
> - mm.1: [76–77] -

> Ah, my daughter, er, like...very much like the zoo
>
> - mm.1: [688] -

The same is true of Phuc:

> Um...don('t) thin(k) too mush...if you thin(k) too mu(ch)...you go feel bore(d) for sure, you know?
>
> - phuc.1: [151–2] -

> Yes...but not much
>
> - phuc.1: [378] -

> he kept talkin(g) and I didn't understan(d) him much
>
> - phuc.1: [582] -

Canh and Long provide examples of much in interrogatives as well as the other permissible structures mentioned above:

> them they very...suffer...ah, I worry very mu(ch)
>
> - canh.1: [1586] -

> Dictator...the people doesn('t) like him mu(ch)
>
> - canh.1: [1173] -

> And even, ah, people live in countryside, they don't know how mu(ch)
> ...h...how...mu(ch) to...er, contact
>
> - phuc.1: [1478–9] -

Oh well, less [?] they talkin(g) very much

- long.1: [443] -

if not overtime, you know, not too much money

- long.1: [356] -

I ask they, ah, repair that('s) how much money

- long.1: [652] -

Of the other informants who use a lot, Ewa S. does not provide evidence of sensitivity to the constraint on much, which she uses as an unadorned quantifier:

kommt vie... much tourist... tourist

- es.1: [427] -

Tam presents a somewhat more interesting case. All his tokens of much are bound up with too:

You know... they fut in too mu(ch) oi(l)

- tam.1: [717] -

In the event, this token and the other five he produces occur in semantically acceptable contexts, where excess actually appears to be involved. Given the fact that he does not produce tokens of much, and that his form of many is an invariant too many, this may be somewhat fortuitous; it is difficult from such evidence as there is to determine whether Tam is producing much only in pre-adverbial contexts, or whether he is simply sticking to a formula. If the latter is the case, however, it may still constitute evidence of sensitivity to the much constraint—the possible semantic non-standardness resulting from invariant use of too much is considerably less striking than the syntactic non-standardness produced by using much in the contexts forbidden by the constraint.

Use of the apparently invariant forms too much and too many is a frequently commented-on feature of learner language. It is possible that use of such forms is conditioned in some way by the much constraint. This could happen in the following way. The learner first encounters or registers much in affirmative sentences, where its lexical shape is always too much (or very much). (On the production side, this is borne out for the Vietnamese speakers, whose first tokens of much are post-adverbial). This initial hypothesis

receives no further checking against structures other than the affirmative sentences where it was first noted, and the result is that the misconstrued lexical item, too much, is used everywhere. The fact that this strategy does not result in drastically ill-formed structures when the item is imported to negatives and questions allows the learner to persist with his or her first hypothesis, which may even receive some confirmation in that post-adverbial much does actually occur, under certain conditions, in such structures.

In the case of the remaining informants who use a lot, there is possible evidence of another kind that they have some sensitivity to the constraint. These informants, Jan R., Krystyna A. and Ludwiga J., simply do not use much at all. As can be seen from Table 4.66, this results in out-of-sequence empty cells on the much row. This is particularly unexpected in the case of Ludwiga J.—that is, without an explanation like the one above.

Infrequently Produced Quantifiers

Finally, certain common quantifiers are notable for their limited distribution. These are each, both and most, which, with the exception of one case of each, seem to be restricted to informants above 1+ on the *ASLPR*, and therefore probably constitute a useful indicator as to oral proficiency, if they happen to be produced. One other quantifier, few, is somewhat more common, but is not found in the output of speakers below level 1 on the *ASLPR*. The narrow distribution and low frequency of each can probably be explained by the fact that the adverbially derived every, a close semantic relation, has wide distribution.

4.14.4 Summary—Quantifiers

To recap briefly, then. Quantifiers, where this is possible, probably emerge from adverbial phrases, or bound adverbial forms. The more distributive quantifiers, such as all and every, and no are relatively more frequently produced by speakers in the lower levels of the *ASLPR*; amongst the less distributive quantifiers, such as many, much and some, various developmental trends can be observed as one progresses up the *ASLPR*. With the article-like quantifiers any and some, language-specific patterns may be reflected, which may lend support to the differential principle of transfer/interference discussed in connection with articles in 4.13

Numbers

4.14.5 Distribution of Numbers

Given that numbers represent a superset of the quantifiers of English, we will finish the discussion of quantifiers with some observations on their production and distribution.

Basically, the tables are all fairly similar, so that the overall frequencies for the four interview groups provide an adequate description of the frequency patterns for numbers for 0+ onwards on the *ASLPR*.

Cardinal Numbers

Amongst the cardinal numbers, one to ten are consistently amongst the twenty most frequent numbers. Of the multiples of ten, hundred and thousand are consistently common, as are twenty and fifty. Zero is generally the form chosen to represent nought.

Ordinal Numbers

Of the ordinal numbers, first is by far the most common. Others seem quite rare.

4.14.6 Summary—Numbers

Numbers are frequent in the output of all informants. Even the lowest ranked informant on the *ASLPR* in the first round of interviews, Van, produces one, two, three and ten.

Clearly, numbers are easy to learn. Their semantics are unambiguous and universal, unlike, say, those of other sets, such as colours. Because of this they can presumably be taught very much as needed. All things being equal, learners would be likely to most need the numbers noted above.

In regard to numbers, one other point is worth making. It is frequently argued that in the area of lexis, there is very little predictability. The patterns described for numbers suggest that this claim may be in need of some qualification.

4.15 Deictics and Demonstratives

4.15.1 Distribution of Demonstrative Pronouns

Tables 4.70 to 4.70 provide figures for the distribution and frequencies of deictics and demonstrative pronouns. These two classes of words are grouped together because for both of them the semantic distinction between proximity and distance, be it spatial, temporal or referential, is marked.

4.15.2 Significant Differences—'This' and 'That'

One notable feature in the pattern of use in these words is that this and here are considerably more widely distributed than that and there.

Deictics and demonstratives share the function of pointing, either literally, across space (That one, over there), or metaphorically, across time (It