*G. Leitner**

Begin and *start* in British, American and Indian English

1. Introduction

'Beginning' verbs are high-frequency words, they share semantic and syntactic properties and occur in a large variety of surface syntactic patterns. They have raised questions in connection with the structure of the complex verb phrase, clause embedding, and on what grounds and how surface patterns can be reduced to one or a small number of more basic structures. Other questions raised are: do such parameters as the meaning of syntactic constructions, selection restrictions that the verbs impose on subjects and object NPs, including the embedded clauses, and the semantics of head nouns in surface subject and/or object position play a role in predicting their syntactic behaviour? When and under what circumstances do speakers talk about the initial phase of some activity. What is the "normal" textual mode of referring to that phase? Is it the narrative or some other pattern? Are there differences relating to text types? And last but not least, are there reflections of potential "epicentres" of English world-wide, which are tied up with individual lexemes, with lexico-grammar, or with broad grammatical systems (Leitner 1992)?

In this paper I will give a progress report on the syntax of *begin* and *start* in the British English LOB, the American English BROWN (BRN), and the Indian English KOLHAPUR (KOL) corpora and argue for considerable differences related to their exploitation of gramatical possibilities, text types and corpora. I will suggest a line of thinking that understands 'fluid' frequency differences as emanating from some more general notional properties.

* *G. Leitner*
Freie Universität Berlin
Institut für Englische Philologie
14195 Berlin (D)

2. Syntax and semantics of aspectual verbs

Though there are only few specific studies of these verbs, a considerable amount of research relates to them. Before I survey earlier studies I will outline the major patterns in which they occur so that the different approaches to, and questions on, the micro-syntax can be better understood.

2.1. The surface grammar of *begin* and *start*

The verbs occur in transitive, intransitive, and link verb functions. The following examples illustrate them:

(1a) *The conductor began/started the concert*
(1b) *John began/started his speech/apology*
(2) *John began/started to run a mile*
(3) *John began/started running a mile*
(4a) *John?*began/started the race*
(4b) *John?*began/started them racing*
(5a) *John began/started by playing the guitar/at 5 p.m./on the podium*
(5b) *John began/started with a guitar recital/at 5 p.m./on the podium*
(5c) *John began/started as a guitar performer*
(6a) *The concert began/started (with a guitar recital)*
(6b) *Mount Everest begins/starts here*
(7) *John began*
(8) *John began/started young/a young man/as a young man*
(9) *"It's a bloody useless job," John began/*started*
(10) *John *began/started for the door*
(11) *John and his girl-friend *began/started (up) a family*

In (1) to (4) *begin* and *start* are used transitively. In (1), there is an NP object, in (2) and (3) a non-finite complement *to* and *ing* clause. (4a) is ambiguous. It may be like (1), meaning that John took part in the race (*begin* is possible). But there may be a causative interpretation according to which John started the race without himself taking part. *Begin* is not usable (but Dixon 1991:177). (4b) is unambiguously causative. (5), (6) and (7) illustrate intransitive uses. (5a/b/c) show combinations with an adverbial of manner, place, time, etc. (6a/b) are instances of intransitive uses without an expected adverbial complement. They differ in that one might conceivably paraphrase (6a) as "Someone started the

concert" while (6b) does not permit such a paraphrase. (7) is intransitive and must be understood as referring to some (ellipted) action/object, as in examples (1) and (2/3). In (8) the verbs are link verbs with an adjectival, nominal or prepositional complement phrase. Finally, (9) to (11) exemplify transitive and intransitive uses where the two verbs cannot be interchanged.

2.2. The search for a unified approach

With such a diversity of surface patterns attempts have been made to suggest more insightful analyses. While early transformational research was confined to the formulation of transformations to account for complement clause embedding in subject and/or object position (Perlmutter 1979), recent research has taken a broader view. The possible interaction of syntax with semantics has been at the centre of Dixon (1991) and Langacker (1991). Lexical semantics were looked at by Leitner (1993), Legler (1975), Schmid (1993) and Lipka/Schmid (1994), and frequency patterns by Leitner (1992), Kjellmer (1992) and Halliday (1993).

2.2.1. Relating the semantics and syntax of aspectual verbs

Dixon (1991) makes several points. Firstly, as these verbs do not introduce an extra semantic role, they belong into the same category as modals, semimodals, 'trying' and 'daring' verbs. All of them (can or must) combine with another verb, and none permits or requires an independent role in the (complement) verb phrase. There is subject identity between the matrix and the embedded clauses and, if the verb is used transitively, object identity (cf. 1a/b). In examples (2/3) *John* is subject of both the 'beginning' and the 'running' action. The concert in (1) is the object of *begin* and would also be the object of a sentence like "John began to open the concert". The exception to this identity constraint is (4b), which might be continued with "(John started the race) to get under way". This and their behaviour in passivization suggests that they form a syntactic and semantic unit with "the verb [they] modify." (Dixon 1991:178, 311).[1]

---

[1] They do not allow passivization of the matrix (i.e. `beginning') clause alone, cf. *The prisoners were begun to count/counting". Passivization of the complement clause alone or together with the matrix clause is possible, *viz.* "They began to be counted/?being counted" and "They were begun to be counted/?being counted". The acceptability judgements are mine.

Secondly, he argues for a unitary analysis of examples (1) to (9) (with the exception of causative *start* in (4) and intransitive *begin/start* in (6b)) by postulating an underlying clausal complement phrase. While it appears on the surface in (2/3), it has been deleted in (1), non-causative (4b), (5), (6a), (7), (8), and (9). In (7) the complement verb phrase has been deleted along with the embedded object NP. Causative *start* can be analysed with a complement clause but does not of course have object identity between matrix and embedded clause.

Just like the causative (4) examples like (6b) are exceptions. They are considered true intransitives and possible only with a limited range of nouns. (1b), which has been called ergative by Sinclair and Halliday, is treated as object-to-subject raising.

Thirdly, the deletion of complement VPs and embedded object NPs is subject to pragmatic, semantic and syntactic constraints. As for the first, retrieval must be possible within context. Dixon claims that only a small range of semantic noun classes are eligible for deletion. For instance, event nouns like *concert*, (1a) and speech act nouns as in (1b) are cases in point. True intransitive uses are restricted to spatial nouns like *Mount Everest*, (4b). One may add "serial nouns" like *alphabet*, *week*, *Tuesday*. Syntactic constraints forbid the deletion of three-place verbs like *give* or *tell* (someone a story). Thus, "John started to give bribes to the director" cannot be reduced to "John started (bribes) to the director". Only two-place verbs like *open, prepare* can be deleted. Such verbs are monotransitives or intransitives at the surface. Moreover, he assumes that speech act nouns (1b) or action nouns (4) could be understood as nominalizations of complement clauses, such as (1b') John began to apologize/speak) (1991:174).[2]

Langacker (1991), on the other hand is concerned with the differences between (1a/b) and (2/3). He argues that in (2/3) the verb profiles a process in which the trajector (= subject nominal) "is a thing (...) and the primary landmark is another process" (1991:197). (1 a/b), on the other hand, "profile a process in which both participants are things" (1991:197) and adds:

---

[2] Of course, one might argue that (1b) should be derived via predicate deletion from (1b'') John started to make/making an apology/speech since a sentence like (i) John started his farewell can only come from (ii) John started to make a farewell (address) where predicate deletion and compound reduction apply.

> *"The initiated process remains as a privotal facet of the base - it's the active of the landmark with respect to the inceptive process designated by this predication - but it need not be spelled out explicitly when its character is apparent from the context or the other lexical items in the sentence." (1991:197)*

In other words, the syntax of *begin* and *start* is based on *alternative* ways of complementation, *viz.* nominal and/or verbal. Whether he would treat intransitive uses as (more) nominal or verbal remains to be seen. But it is clear that the syntax would not be reduced to one dominant pattern. He assumes a semantic difference between nominal and clausal objects.

Dixon (1991), Quirk *et al.* (1995) and others also looked at potential differences between the two complement types, *viz. ing* and *to*-infinitive. There is agreement that, despite considerable semantic neutralization, a *to*-clause indicates potentiality, an *-ing* clause actual involvement. Quirk *et al.* argue that the latter is preferred if iterative actions are described. And Dixon relates the choice to the meaning of the syntax of "(*for*) *to*" clauses on the one hand and "modal *-ing*" on the other. The former implies the potentiality of the subject's involvement, the latter actual involvement. Both assume that the choice is absolutely free ("null hypothesis") so that the verbs could use either constructions.

The semantic difference is slight and often neutralized as in (2) and (3). To the extent that it exists, *ing*-clauses would relate 'beginning' verbs to achievement verbs as in

(12) John managed to go to the concert ('=' John went ...)

(13) John made it to the concert ('=' John was at ...)

Similarly nominal objects, (1a/b), (4a) or

(14) John started a family/a company ('=' a family/a company existed)

(15) "That's absurd," John began ('=' these words were said)

side with actual involvement and imply achievement. (14) could not be continued with "but then decided otherwise", nor (15) with "but he didn't come to the word 'absurd'".

2.2.2. Lexical semantics of *begin* and *start*

As for lexical semantic differences Lipka/Schmid (1994) and Schmid (1993) argue that *begin* is more usual in non-agentive and ergative con-

texts like (6a), *start* in "canonical agentive" constructions like (2) and (3). More importantly, *begin* signals "gradual beginnings, mostly in the cognitive or emotive domains, and for contexts involving speaking and talking" (1994:13), *start* "is used to denote dynamic and sudden beginnings of actions" (1994:13). They conclude that the semantics of *start* is more complex than that of *begin*, which is confined to the meaning "inchoative".

2.2.3. Corpus evidence

Halliday (1993) used the Cobuild Corpus to investigate the assumption that syntactic information is not patterned arbitrarily but follows two radically different patterns, *viz*. 0.5.:0.5 and 0.9:0.1. In order to test this "equi vs skewed probability hypothesis" he looked at polarity and (primary) tense. He found that polarity follows the 0.9:0.1 pattern, tense the 0.5.:0.5 one. In other words, out of ten sentences, nine would be positive, one negative; but about half would be present, the other past. He lists individual tensed verbs which show that not all of them follow that equiprobability pattern. *Begin*, for instance, is clearly out of line (Halliday 1993:19), being past-oriented.

Kjellmer (1992) looked at the frequency of passivity, tense, and *to*-complementation with high frequency verbs in the LOB corpus. He was able to distinguish verbs with low and high present tense orientation, *to* complementation and passivity. *Begin* figures in the low-passivity list (Kjellmer 1992:338) but is not amongst the thirty most frequent low past tense verbs. Nor does it surface with any marked preference for *to*-complementation.

Leitner (1993) compared the description of *begin* and *start* in three major English learners' dictionaries from the point of view of completeness of coverage of data, descriptive adequacy, and exemplification of use. He suggested that a semantically-based approach to syntax (Dixon 1991) may well help reduce the number of sense distinctions (or, in Sinclair's model, sense-grammar pairings), lead to a deeper understanding and to easier information access for dictionary users. Drawing on the LOB corpus he also noticed certain common deficiencies. As the two verbs showed considerable frequency differences in syntax, the question must be raised how relative frequencies differences should be accommodated in dictionaries and grammars.

Corpus studies then have suggested certain preferences in choice of complementation, tense use etc. which are worth studying and which cast considerable doubt on the "null hypothesis". They provide a challenge to see if, and how, frequency differences may be related to semantic-syntactic theories.

3. *Begin* and *start* in Brown, LOB and Kolhapur

The following analysis is concerned with micro-syntax. But as several issues have to do with general syntax, the findings will have to be interpreted in a broader context so as to see what is characteristic of the verbs proper and what is merely an instantiation of other systems.

3.1. Method

All tokens of *begin* and *start* were retrieved from the three corpora along with complete sentence context which was considered preferable to concordances for the future study of the semantic parameters that may bear upon syntax. Four systems at phrase and clause level were identified and tagged (Quirk *et al.* 1985):

- subject (encoded as SJ),
- verb phrase (VB),
- nominal complement or non-finite complement clause (TR),
- adverbial complements (AC).

Tagging being manual, only VB- and TR-systems have been completed.

3.2. Findings

The findings will show considerable imbalances in the syntactic behaviour of the two verbs and that the assumption made earlier that there is a latent tension between syntax and semantics.

3.2.1. An overall view: written and spoken English

The corpus consists of 1520 tokens of *begin* and of 1121 tokens of *start*. Occurrences of the non-finite adverbial phrase *to begin*/*start with* have been counted separately as was the one nominal subject clause in which *start* occurs.

Table 1: Tokens of *begin* and *start* in the corpora

The two verbs account for 0.088 % (*begin* 0.050 %, *start* 0.037 %) in the 3-million word corpus. *Begin* is more frequent than *start* in each corpus of written and printed speech. But this pattern is reversed if one compares it with the spoken LLC corpus.

3.2.2. The Verb Phrase

The term VB is used in the narrow sense of referring only to the verbs (Quirk *et al.* 1985, ch. 3).

3.2.2.1. Finite and non-finite VPs

Both verbs occur in non-finite clauses and more frequently in finite ones, and the imperative mood, tables 2 and 3:

Table 2: Finite and non-finite verb phrases and systems operating at non-finite VPs

Begin and *start* behave quite differently in this system. 19.7 % of all tokens of *start* occur in non-finite clauses as against 7.4 % of *begin*. *Begin* is barely used in the imperative mood.

Table 3: Non-finiteness split up into systems

Table 3, however, reveals similarities within the non-finiteness system. *To*/ø-constructions come second to participial clauses for both. The former accounts for 41.6 % of *begin* and 43.9 % of *start*, the latter for 58.4 % and 54.7 % respectively. 48.7 % of *begin* and 44.8 % for *start* use *-ing* clauses. Participial *-ed* constructions are third with under 10 % each.

In other words, the verbs exploit (non-)finiteness to different absolute degrees but inside non-finiteness they are quite similar.

3.2.2.2. The simple finite VP

At the simple finite verb phrase tense is the only system and distinguishes (primary) present and (simple) past tense. Although present tense morphology need not necessarily denote present time, that semantic distinction was not investigated. Present tense indicators are the use of the base form as a finite form, the third person *-s*-form, and the *do*-periphrasis in negation, emphasis and inversion (in questions, after negative adverbs etc.). Past tense indicators are the finite past *-ed*-form and likewise the *did*-periphrasis. Table 4 has the details:

Table 4: Systems operating at the simple VP

It appears *prima facia* that the two verbs differ radically with regard to tense orientation, with *begin* being far more past oriented than *start*. This impression must, however, be modified somewhat in the light of the finite/non-finite VP findings since the relevant corpus for tense orientation is not the total corpus, *viz*. 1520 and 1121, but the number that remains after non-finite VPs and imperatives (see line 4, table 2) and complex finite VPs (line 1, table 5) have been subtracted.

But a sizeable difference remains: 17.0 % of *begin* and 25.0 % of *start* are present tense oriented. Both verbs are quite skewed compared with regard to Halliday's patterns.

3.2.2.3. The complex finite VP

At the complex VP operate the systems of modality (A), perfect (B), progressive (C), and passive (D). Four degrees of complexity can be distinguished, *viz*. singular, dual, triple, and quadruple complexity.

They depend on the number of systems which are simultaneously selected in a given VP. Only dual complexity, the combination of two systems, was found. Table 5 has the results:

Table 5: Summary of systems operating at complex finite VP

3 This includes one *see* + *ing* construction.

Several results are noteworthy. Firstly, 26.6 % of all VPs are complex (see 3rd col., table 4). In terms of the preference of the systems, perfect precedes modality, progressive, and passive. Secondly, there are subtler differences between the two verbs. For instance, *start* occurs more frequently in complex VPs than *begin*. 31.5 % of all tokens of *start* in finite VPs are found in complex VPs, as against *begin*'s 23.6 %. *Start* is more frequent in dual complex VPs. 3.6 % of *all* VPs of *start* or 11.3 % of *complex* VPs occur in dual complex VPs, while *begin* hardly occurs at all. Finally, the verbs prefer different systems: *start* favours system A more than *begin* does, *viz.* 38.3 % vs 27.1 %. It is particularly frequent with D, with 11.6 % of all complex VPs. Adding up all occurrences of passives, irrespective of the complexity of the VP, *start* is used 56 times, i.e. in 20.4 % of all cases, *begin* in 4.5 %. In contrast, *begin* is more frequent in C, with 28.0 %, as against 3.9 %. These findings are accentuated if the verbs are compared *within* the four systems. Thus, *start* accounts for 53.8 % of combinations with modals, for 71.1 % of passives (or 80.0 % as for D, AD, BD and CD). *Begin* stands out with system B, the perfect, (56.9 %), and C, the progressive (91.2 %).

3.2.2.4. Mood

Let me repeat here that the only feature of interest was the occurrence of *start* in the imperative cf. line 3, table 2.

3.2.3. Transitivity

Within transitivity the choices of complementation were studied, i.e. whether they occurred without a complement (ø-comp), a nominal direct object complement (NP-comp), or a non-finite *to*-infinitive (*to*-comp) or *ing*-clause (*ing* comp). The frequencies are in table 6:

Table 6: Frequencies of *begin* and *start* by transitivity

Nominal complementation (i.e. ø-comp and NP-comp) contrasts with verbal complementation (Dixon 1991; Langacker 1991). There are striking differences which need to be sorted in different ways.

Start is particularly strong with nominal complementation (66.5 %), *begin* with verbal complementation (58.9 %). Within these patterns *start* prefers ø-comp over NP-comp somewhat more (difference of 20.7 %) than *begin* does (difference of 17.1 %). But in regard to verbal complementation types, the patterns are reversed *Start* favours *ing* over *to*-comp, *begin to-* over *ing*-comp.

To conclude with interaction of TR with complex VPs. This shows a definite accentuation of these preferences, see table 7.

Table 7: Complex VPs by TR

Start here favours ø-comp even more than overall (48.2 %), but levels out with the other types. *Begin* is even stronger with *to*-comp (60.2 %) and considerably weaker with *ing*-comp (3.6. %).

3.2.3.1. Simple VP: tense

Looking at the simple VP it appears that the differences are particularly pronounced with ‘ø-comp’ and ‘*to*-comp’.

Table 8: Simple VP (tense orientation) and TR-types

It seems clear that the basic pattern with *start* being more 'nominal' and *begin* being more 'verbal' remains. However, there are divergent trends as one moves from present to past tense. A comparably higher number of present tense *begin* and *start* use ø-comp or NP-comp, *viz.* 58.5 % and 65.8 %, respectively. In the past, these types decrease, but *begin* drops to a low 33.4 %, *start* to only 56.5 %. Conversely, 'verbal' complementation rises to 66.6 % with *begin* and to 43.5 % with *start*. Notice here that the use of ing-comp with *begin* is almost entirely due to past tense, a pattern that is not too pronounced with *start*.

Past tense also minimizes the use of ø-comp, *viz.* 20.3 % for *begin* only, and 35.9 % for *start*. And there is a marked rise in *to*-comp with *begin* to 54.0 % and with *ing*-comp to 12.6 %. Past reports of 'beginning' actions tend to be more verbal generally, but the differences between *begin* and *start* are only accentuated.

3.2.3.2. Complex VP: other systems and combinations

System B, perfective, reveals a preference for verbal complementation, which is more marked with *begin*.

Table 9: System A, perfective, and TR

System C, progressive, bolsters up the finding with tense orientation. It favours a verbal complement type definitely for *begin*. For *start* little can be said beyond the suspicion that all patterns seem to be used.

Table 10: System C, progressive and TR

System D, passive, shows a unique, categorical, preference for nominal expression and intransitive use at that.

Table 11: System D, passive, and TR

While 50 % each of *start* occur in ø/NP-comp and *to/ing* comp respectively, the pattern is skewed for *begin* in favour of the latter type (62.2 %). Once again, *to*-comp is more frequent with *begin* (54.8 %) and *ing*-comp with start (31.4 %).

3.2.4. Summary of VP and TR

Start has a more balanced overall syntax than *begin* if one takes into account (non-)finiteness, simple/complex VPs and (to a lesser degree) tense orientation. As for complex VPs, *start* is more passive-oriented and favours complex dual VPs. *Begin* prefers singular complex VPs and the progressive. It is also not used in the imperative, where *start* seems to be preferred.

It has been shown that the assumption of an identical exploitation of the systems operating at VP and TR cannot be maintained. It is of interest now to see if and how these findings correlate with text type and varieties.

3.2.5. Text types and varieties: overview

Text type and varieties bring in elements from outside the language system proper, *viz.* the norms governing the construction of texts and of communication at large. Various patterns were cross-tabulated. Given the disbalances between non-fiction and fiction categories (about 3/4 are non-fiction) some of the findings were made to correspond to a 1-million word corpus for each. See table 12 below.

Table 12: *Begin* and *start* for text categories (non-)fiction

The three corpora do not use *begin* and *start* in the same ways. Line 7 suggests a much higher frequency of *begin* in BRN, which disappears if one looks at the corrected data in line 8. The corrections are, of course, the effect of the relative frequency differences between +F and -F, i.e. categories K-R and A-J. The greater the difference the bigger the balancing effect of the corrections. As for *start*, the effect is particularly strong with LOB, which now comes out first.

The normalized data show that in balanced corpora the preference for *begin* and *start* in non-fictional texts would be minimized. The two verbs would in fact be much more frequent in fictional categories. In LOB, *start* would by far dominate +F, while *begin* would be more frequent in BRN and KOL.

3.2.5.1. Text types and the -F/+F distinction

As for individual textual categories a frequency ranking leads to the observation that categories F, G, and K occur amongst the five most frequent ones with both verbs, J and L occur with *begin*, and E and A with *start*, cf. table 13.

Table 13: *begin* and *start* by most frequent text categories

While the first five categories account for around 2/3 of the tokens of *begin*, it is only just over half with *start*. I take this to imply a fair amount of overlap but would suggest, firstly, that it may ultimately not be the -F/+F dimension that is relevant but some other text linguistic criterion, and, secondly, that there may be subtle corpus-related differences.

The observation, that *start* seems to occur more frequent in -F-(and the five categories in particular), can be established on firmer grounds. 57.6 % of *begin* but 78.8 % of *start* are in those five categories. If one excludes the popular literary category G and corrects the -F-figures accordingly, the percentage drop to 42.6 % for *begin* but to only 71.8 % for *start*.

The following systems were found to be of interest here. In the -F category there are 463 occurrences of past tense *begin* and 428 in +F. If the data are normalized to 1 million each, one realizes that about 1/3 of past tense occur in -F and 2/3 in +F. As for *start* a different picture emerges: past tense would only be somethat higher in +F.

As *start* is more present oriented, the question arises whether there are peak categories in -F. In fact there are. Of a total of 149 tokens nearly one hundred occur in category A, E, F, G, and J. Tense choice then strongly correlates with text type not necessarily with -F/+F.

It seems rather that we have text type and text specific constraints that require further investigation. To just mention two observations. G ("Belles Lettres") with its historical texts contrasts with political reports (category A), science descriptions (category I) etc. A look at occurrences in individuals texts yields a considerable number with more than three tokens (Brown E 26 has over ten even). One suspects that the occurrence must be seen as a reflection of particular themes rather than (directly) of text types.

3.2.5.2. The corpora

As with text categories, one must not be blinded by the search for differences. There is a substantial degree of overlap. For instance, table 12 revealed the preference of those verbs in +F and the (simple) past at that. Also, five out of nine -F categories accounted for a large percentage of tokens in all corpora. And all syntactic patterns, (1) to (11), are used.

But as with text categories, I will explore the possibility of differences which may reside in two areas. They may either be found in the lexico-grammar of the verbs or the broad grammatical systems and their meanings in which they take part. Evidence for the latter would be established if broad systems were found to behave analogically across corpora. The lack of that would point to lexeme-specific differences. As no comprehensive study could yet be made, I will concentrate on some illustrative areas.

3.2.5.2.1. *Start*

To begin with simple finite VPs and TR (table 6). See table 14:

Table 14: *start* in simple VPs by TRs

Clearly, KOL's share of simple VPs is somewhat higher (41.5 %) than its overall share (see table 12 with 37.7 %). As for TRs it accounts for 59.1 % of all *ing* and 40.6 % of ø-comp. It definitely underuses *to*-clauses (16.9 %). If one looks at *start* in simple VPs inside KOL, there are two clear peaks, *viz.* ø (38.3 %) and *ing* (40.7 %). *To*-clauses only account for 5.2. % of KOL's tokens.

The two other corpora are somewhat more balanced, i.e. *start* uses the four TR possibilities more evenly. Overall, ø-comp is the first choice in all corpora (LOB 41.8 %), BRN (39.1 %), along with KOL's (38.3 %). *Ing* and NP are almost equally strong in LOB and BRN. A clear difference only occurs with *to*-clauses with BRN accounting for 52 % of all tokens.

A look at table 15 reveals more subtle patterns still.

Table 15: *start* in s. present and past by TRs, -F/+F and corpora

It is clear that the three corpora are ordered with regard to present and past orientation. Thus, LOB is most (27.9 %), KOL next (24.6 %) and BRN least (21.2 %) present-oriented (see 2/b, c, d by the sums of col 2 and 8). But if one considers the corpora's contribution to present or past tense choice, it emerges that KOL's disproportial use of start is equally balanced in the two tenses (around 41 %). BRN on the other hand is least present, LOB least past oriented (around 26 % respectively).

To turn to tense and TR choice. The preference of KOL for *ing*-comp has been established above. Again, this is unbiassed with regard to tense choice (i.e. around 40 % each). But its contribution to the overall effect is somewhat higher in the past (63.1 %) than the present (57.9 %) (see 4/c by 2/c and 10/c by 8/c). As for BRN and LOB, it seems that the former comes second in present tense choice (21 %) and LOB last (13 %). No other differences emerge.

To-comp was seen to be particularly weak with KOL. But its use rises somewhat, again in the general pattern, in the past to as much as 17.4 % (see 9/c by 9/e). BRN and LOB differ in the former being much more present-, LOB most past-oriented with *to*-comp (around 57 % each). Nominal complementation, favoured with *start* generally, shows much the same effects. There is a considerable drop with BRN from 68.4 % in the present to 53.9 % in the past (see 3, 6/b by 2b and 8, 11/b by 12/b). These shifts are less pronounced with KOL and LOB.

3.2.5.2.2. *Begin*

A brief comparison with *begin*, table 16, confirms the differences already described, in particular its slightly lower present tense orientation, the sparcity of *ing*-comp in the present and the huge shift to verbal expression in the past.

Table 16: *begin* by a present and past tense, TR, and corpora

It can now be seen that these patterns are also the results of corpora influences. Thus, the rise in *ing*-clauses is least in LOB with only to 6,9 % in the past (see d/4 and d/10). It is particularly strong with KOL's increase to 16.2 % (c/4 and c/10).

3.2.5.3. The roles of text types and corpora

There is no scope for further analyses of the role of text types and corpora (Leitner, in preparation). But the evidence presented suggests clear influences related to the parameter of text categories. *Start* definitely occurs less in fictional texts and, given its relative preponderance in the present tense and more frequent use in non-finite clauses systems, it may be said to also favour a more colloquial and/or factual style of writing. That inference is reinforced by a cursory look at the spoken London-Lund Corpus which shows it to be about twice as frequent as *begin*.

Irrespective of the welter of overlapping properties, corpora yield some specific patterns. Moreover, it is impossible to contrast first lan-

guage varieties as being more similar to each other with second language varieties, represented by KOL, or even to eliminate their effects as insignificant altogether.

4. Discussion: Locating and unifying the findings

Summing up his findings on passivity, tense and complementation, Kjellmer refers to "variations and exceptions within some rule-defined domains of the English verb ... which cannot all be accounted for by reference to some general principle." (1992:342). He adds that these are "lexically based restrictions" and echoes Algeo's earlier evaluation of British and American differences:

> *"... no general conclusions come leaping out of the mass of data under inspection - indeed, it is my assumption that there are no general patterns of difference .... It appears as though most of the grammatical differences are not syntactic generalizations that might appear in a grammar book, but rather are lexical features that one would expect in a dictionary." (1988:3)*

Halliday (1993) was not concerned with the behaviour of individual lexemes within the systems of polarity and tense. But, while he saw his assumption confirmed that two radically different grammatical patterns of information distribution would be established, these generalizations were hiding significant lexically-based differences. The differences established between *begin* and *start* between themselves and across text types and corpora make the same fundamental point. If there is a core of English (Leitner 1992), then it must be enriched by a lexicallly-based dimension that leads to frequency patterns.

But can we say more? I have tentatively used notional labels such as action- vs. participant-, present- vs. past-orientation as umbrella terms to unify the findings in the VP and TR systems. These terms derive from Langacker's notion of the profiling of actions/processes and things (1991), as well as from Dixon (1991). *Start*, thus, has a greater tendency of profiling 'things', *begin* of 'events'. Both verbs profile processes more in the past, than in the present. On the assumption that the function of the passive and the ergative is to focus (or profile) 'things', rather than processes, *start* would naturally favour passives and ø-complementation, *begin* the active and verbal complementation.

One might wish to make a theoretical point. Given the imbalances established, Langacker's suggestion of *alternative* ways of comple-

mentation is closer to an adequate understanding of these verbs than Dixon's postulate of *one underlying structure*. And this appears to be borne out if we look at other 'beginning' verbs. Only *commence* allows verbal *ing*-comp, all others, eg. *initiate, introduce, set off, set out, set up, became -ed* participle, are nominal and profile things either as agents (in the active mood) or events. As pointed out, nominal orientation relates the verbs to achievement verbs. The actual realization of what the object NP refers to is presupposed. And if we add that *ing*-comp implies actual involvement and that a *to*-clause, given choice, reflect mere potentiality, we could establish a lexical-semantic cline from profiling things to processes. While both *begin* and *start* can use all options on that cline, *begin* is more at the 'process' end, *start* more in the middle. *Commence* would be further to the 'thing' end and *set off, initiate* etc. can only profile 'things'. In other words, the way we can talk about the beginning phase is tied up with general semantic categories and a structured lexical field.[4]

References

Algeo, J. (1988): British and American grammatical differences. In: *International Journal of Lexicography*, 1(1). 1-31.

Dixon, R.M.W. 1(991): *A new approach to English grammar, on semantic principles.* Oxford: Clarendon Press.

Halliday, M.A.K. (1993): Quantitiative studies and probabilities in grammar, in: Michael Hoey (Ed.). In: *Data, description, discourse. Papers on the English language in honour of John Mc Sinclair*. London: HarperCollins, 1993, 1-25.

Kjellmer, G. (1992): Grammatical or native-like?, in: G. Leitner, ed., *New directions in English language corpora*. Berlin: Mouton-de Gruyter. 329-344.

Langacker, R. (1991): *Concept, image, and symbol. The cognitive basis of grammars*. Berlin: Mouton-de Gruyter.

Legler, B. (1975): *Infinitiv und Gerundium nach den Verben BEGIN, CONTINUE, CEASE (etc.) in der englischen Allgemein- und Fachsprache der Gegenwart*. Dissertation an der Philosophischen Fakultät der Universität Halle-Wittenberg.

---

4 Questions remain. I have been unable to suggest a solution for the fact that *start* is more frequent in the imperative and in subordinate non-finite VPs. And Lipka/Schmid's assumption (1994) that *start* carries much of its `etymological' meaning and is more often used for abrupt, *begin* for slow beginnings needs further study. It is as yet unclear how frequency patterns correlate with such a (connotative?) notion.

Leitner, G. (1992): English as a Pluricentric Language, in: M. Clyne (ed.), *Pluricentric Languages*. Berlin: Mouton-de Gruyter. 178-237.

Leitner, G. (1993): "Where to *begin* or *start*? Aspectual Verbs in Dictionaries". In: Michael Hoey (Hsg.), in: *Data, description, discourse. Papers on the English language in honour of John Mc Sinclair*. London: HarperCollins, 1993, 50-63.

Leitner, G. (in preparation): Aspectual verbs of beginning: a corpus-based comparative study.

Lipka, L., H.-J. Schmid (1994): *To begin with*: degrees of idiomaticity, textual functions and pragmatic exploitations of a fixed expression. In: *Zeitschrift für Anglistik und Amerikanistik* 42(1). 6-15.

Perlmutter, D. (1979): The two verbs *begin*. In: R. Jacobs, P. Rosenbaum, eds., *Reading in English transformational grammar*. Waltham, Mass.: Ginn and Company. 107-119.

Quirk *et al*. (1985): *A comprehensive grammar of the English language*. London: Longman.

Schmid, H.-J. (1993): *Cottage and Co., idea, start vs. begin*. Tübingen: Niemeyer.