Use of tenses by advanced EFL learners: evidence from an error-tagged computer corpus

Sylviane Granger
Centre for English Corpus Linguistics
Université Catholique de Louvain

in H. Hasselgard & S. Oksefjell (eds.) *Out of Corpora. Studies in Honour of Stig Johansson*, Rodopi: Amsterdam & Atlanta, 1999, 191-202

DRAFT

1. Introduction

One recurrent complaint among teachers of advanced English is the dearth of pedagogical material aimed specifically at this higher level. There seem in fact to be two main reasons for this gap in the market. The first is commercial: the number of learners aiming at a rudimentary or intermediate knowledge of English has traditionally far outnumbered that of learners aiming at advanced or near-native proficiency, and publishers have thus deemed it commercially much more profitable to produce materials targeting the lower proficiency range. The second perhaps related reason is that much less research has been carried out into the upper stages of the second language acquisition continuum and as a result, ELT materials designers do not know exactly what needs to be included in the advanced syllabus.

This situation is certainly changing however, as the status of English as the language of international communication leads to increased demand for advanced levels of spoken and written proficiency. At the same time, a new source of data - the computer learner corpus - has made its way on the ELT scene, creating new opportunities for better understanding advanced learners' needs.

In this article I will demonstrate the potential of 'error tagging', a new type of corpus annotation, for bringing new insights into advanced interlanguage grammar in general and tense usage in particular.

2. Computer learner corpora and error tagging

Computer learner corpora (CLC) are electronic collections of spoken and written texts produced by foreign/second language learners. One of the major advantages of using learner productions in machine-readable form is that the data can be submitted to the wide range of software tools and methods developed by corpus linguists over the last thirty years. Although "like any healthily active and developing field of inquiry, learner corpus research has to continue to face challenges both material and intellectual before it wins a secure and accepted place in the discipline of applied linguistics" (Leech 1998:xx), several studies into learner written and spoken language have demonstrated the value of CLC for second language acquisition studies.[1] In particular, comparisons between native and learner corpora have brought to light some major differences in the use and frequency of some of the key grammatical, lexical or discourse features which contribute to the 'foreign-soundingness' of advanced learners' productions.

While computerization greatly facilitates the retrieval of a variety of linguistic elements in learner data - word partials, words, sequences of words, parts-of-speech, syntactic structures -, it is unfortunately of little use when it comes to detecting errors since current grammar and style checkers can retrieve only a very small proportion of them. However, while the error detection stage is a largely manual one, the subsequent stages of the error analysis process can be partly or fully automated. The Computer-aided Error Analysis (CEA) system developed at Louvain uses this approach. After the initial process of setting up the system, which involved creating an error tag set, writing an error tagging manual and designing an error editor[2], the error analysis system becomes a two-stage process where in the first instance, the errors in a given learner corpus are detected and corrected and assigned appropriate tags (a

[1] See Granger 1998 for a general introduction to the learner corpus field and a series of case studies.
[2] *UCLEE,* the *Université Catholique de Louvain Error Editor,* was developed to speed up the insertion of tags and corrections into the text files thanks to an error tag menu and a correction box. The software was written for the Louvain project by John Hutchinson from the University of Lancaster. It is available from the Centre for English Corpus Linguistics together with the error tagging manual (Dagneaux et al, 1996).

manual process) and in the second instance, the error editor is used to insert error tags and corrections into the text files (a semi-automatic process).

As the system and the principles underlying it have been described in detail in a recently published article (Dagneaux et al 1998), I will restrict myself here to a very brief outline of the approach. The system is descriptive: it tags errors in terms of linguistic categories, not in terms of the source of error. It is hierarchical: it contains 9 major categories and several levels of subcategorization. It is flexible: subcategories can be added or deleted. The nine major categories are: **F**ormal, **G**rammatical, Le**X**ico-grammatical, **L**exical, **R**egister, **W**ord redundant/word missing/word order and **S**tyle. These major category codes are followed by one or more subcodes. For instance, the GVT code refers to grammatical verb tense errors, GVAUX to auxiliary errors, etc.

Once inserted into the learner corpus, the error tags can be searched like any other element in the corpus and comprehensive lists of specific error types can be drawn up. For this investigation, I retrieved all verb tense errors (GVT) from two 75,000-word error tagged learner corpora of argumentative essay writing. The first corpus contains writing by French-speaking university students of English in their second year of study and the second, writing from fourth year students. I will refer to the two corpora as post-intermediate (P-Int) and advanced (Adv) respectively. While the study is not strictly speaking longitudinal since it involves two different groups of learners, it seems legitimate to make inferences in terms of progress as the two groups have followed the same programme of study in the same university.

3. Frequency of verb tense errors

GVT is a subcategory of the GV category, which includes all grammatical errors affecting verbs. The other subcategories are GVAUX (auxiliary errors), GVNF (finite/non-finite errors), GVN (concord errors), GVM (morphology errors) and GVV (voice errors). Table 1 gives the breakdown of the GV category.

Table 1: Breakdown of the GV category

Category	Whole corpus	P-Int corpus	Adv corpus	Progress rate
GVAUX	317	239	78	67%
GVT	248	150	98	35%
GVNF	91	57	34	40%
GVN	72	54	18	66%
GVM	29	21	8	62%
GVV	17	13	4	69%
Total GV	774	534	240	55%

The GVT category is the second largest subcategory of verb errors in the whole 150,000-word learner corpus. It accounts for 32% of the GV errors, immediately behind GVAUX errors, which account for 41%. Following Johansson & Lysvag (1987:117), we have included in the tense category both the tenses proper (present and past) and the forms which are in effect combinations of tense and aspect (present perfect, past continuous, etc.). However, unlike Johansson & Lysvag, we have excluded modal forms (*will/shall*), which in our system are classified as GVAUX .

A comparison of the post-intermediate and advanced corpora revealed that the tense errors displayed the lowest progress rate: an improvement of 35% against the more encouraging 67% for the auxiliary error category. This result, which is extremely disappointing as a great deal of time and effort is spent on tenses in the students' curriculum, leads us to ask a series of questions. Are some tenses faring better than others or is it an across-the-board phenomenon? If - as one might expect - there are differences between the tenses, which tenses are the most error-prone or improvement-resistent? More fundamentally, what can be done to teach these tenses more effectively? The following sections are a preliminary attempt to answer these questions.

4. Error-prone tenses

Table 2 gives the breakdown of the erroneous tenses.[3] The following categories are distinguished: SPr (simple present), SPa (simple past), PPerf (present perfect simple and continuous)[4], PrC (present continuous), PaPerf (past perfect simple)[5] and PaC (past continuous).

Table 2: Breakdown of erroneous tenses

Tense category	N. of tense errors	% of tense errors
SPr	76	32.5%
SPa	68	29%
PPerf	31	13%
PrC	31	13%
PaPerf	18	8%
PaC	10	4.5%
Total GVT	234	100%

The two tenses which provoked the most errors were the simple present and the simple past, although this finding may in fact simply be due to the higher overall frequency of these tenses in English (for a comprehensive survey of tense frequencies in spoken and written English, see Kennedy 1998: 122-230). The raw frequencies of wrongly selected tenses in the whole learner corpus as presented in Table 2 may therefore not prove very revealing. However, a comparison of the frequencies in the post-intermediate and the advanced corpus brings interesting insights into learners' progress. Table 3 for example, shows that as students progress, they make many fewer errors with some tenses, whilst others continue to cause difficulties (compare the 60% improvement with the present continuous with the under 20% improvement rate for the simple past and the present perfect).

Table 3: Breakdown of GVT errors in the two learner corpora

Tense category	P-Int corpus	Adv corpus	Progress rate
PrC	22	9	59%
SPr	48	28	42%
PaPerf	11	7	36%
PaC	6	4	33%
PPerf	17	14	18%
SPa	37	31	16%

For each GVT error, the error-tagged corpus contains not only the erroneous tense, i.e. the misselected tense, but also the target tense, i.e. the tense that should have been selected. Figure 1 shows there to be radically different proportions of erroneous use/ target use across the range of tenses. At the left of the continuum, learners make many errors with the PaC but relatively few failing to select it. At the other end of the continuum, the PPerf very often fails to be selected, with proportionately fewer errors made with actual use. In the middle, it is interesting to note that the two present tenses have similar patterns of misuse and non-use.

[3] For this study only finite verb forms are taken into account. The 14 non-finite errors (such as *I prefer not to live $to have lived$*) are left out.

[4] The continuous forms of the present perfect have been grouped together with the simple forms because they are very infrequent in the corpus (only 3 instances).

[5] The corpus does not contain any erroneous use of the past perfect continuous.

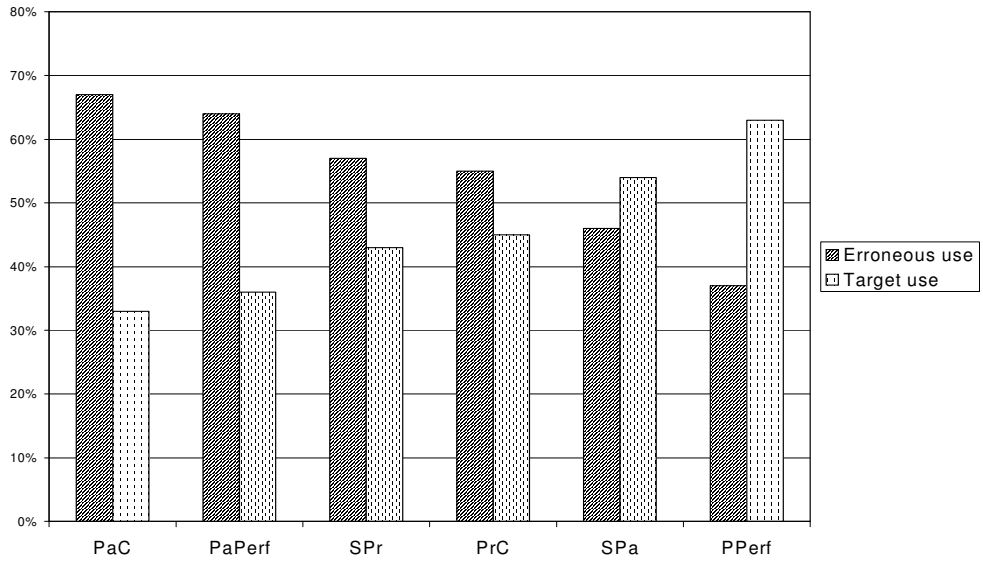

Figure 1: Erroneous vs target use

A detailed analysis of the 'erroneous tense-target tense' pairs brings out the tenses that French-speaking learners are most likely to confuse. The most error-prone tense pairs in the whole learner corpus are listed below in decreasing order of frequency. The figures between brackets indicate the number of instances.

1. Simple Present ⇔ Simple Past (59)
   SPr* ⇒ SPa (38)
   SPa* ⇒ SPr (21)

2. Present Continuous ⇔ Simple Present (53)
   PrC* ⇒ SPr (30)
   SPr* ⇒ PrC (23)

3. Simple Past ⇔ Present Perfect (51)
   SPa* ⇒ PPerf (33)
   PPerf* ⇒ SPa (18)

4. Simple Present ⇔ Present Perfect (20)
   SPr* ⇒ PPerf (13)
   PPerf* ⇒ SPr (7)

5. Past Perfect ⇔ Simple Past (19)
   PaPerf* ⇒ SPa (13)
   SPa* ⇒ PaPerf (6)

In each of the pairs there is one dominant member which accounts for over 60% of the instances. For instance, the figures show that French learners are much more likely to use a simple present in place of a simple past than the opposite. In one case, however, the dominance is less marked: the tendency to use the present continuous instead of the simple present is only slightly greater than the opposite tendency.

The results of this quantitative approach to the data are potentially very useful. They highlight the tenses and tense pairs which should be focused on when teaching French-speaking EFL learners who have reached a relatively high level of proficiency. However, only a minute analysis of the data will provide information on the exact source of the students' problems and ensure that appropriate ELT

tools be developed to cater for their needs.

5. Sources of tense errors

In this section I will describe two factors at the root of a large number of French-speaking learners' tense errors: a sentence-level approach to tense selection and L1-L2 discrepancies.

5.1. Sentence-level approach

A large proportion of tense errors is due to the students' adopting a sentence-level or clause-level approach to tense selection. Examples (1) to (6) are good illustrations of this. Note that all examples in this section are faithful reproductions of the students' texts with only the tense errors highlighted and the target forms enclosed within two dollar signs.

(1) *Since the beginning of the twentieth century*, movements of population have changed the face of our West European towns and countries. A lot of people settled down $have settled down$ with their families and acquired the nationality of our contries. Others are *still* coming.
(2) During two hours I took part to their exercises and animated discussions about tenses in English or the use of conjunctions. The classroom breathes $breathed$ joy and the students were full of life.
(3) What are the different purposes of making music? If the first purpose of music is to entertain, it has now its own power and became $has become$ a big industry which employs millions of people.
(4) Finally the big ethic problem of the abortion. I agree with this practise in some circumstances. For example when a woman was raped $has been raped$. I can understand that it will be difficult if not impossible to keep the baby.
(5) The university is indeed a place where the student learns most of the time only theory about whatever subject he chose $has chosen$, and that causes damage for his professional future: it means that this institution gives him in fact very little opportunities to apply his knowledge in very practical circumstances.
(6) The best way to be sure that a criminal won't act like he did is to put him in jail. But even this isn't a good solution. Some criminals committed $have committed$ such awful crimes that the only sentence can be death.

The sentence containing the tense error in example (1) has two characteristics shared by many other sentences displaying tense errors in the corpus. First, there is no tense error if it is taken out of context. Secondly, it contains no overt time adverbial such as *ago, yet, just, since, for* which are usually presented as triggering the use of particular tenses (simple past with *ago,* present perfect with *yet,* etc.). It is striking to note that in the other two sentences in example (1), which do contain overt markers (respectively *since* and *still*), the correct tense has been chosen. In other words, the student seems to be operating sentence by sentence. In the absence of overt time markers, he selects the tense according to some general principle such as 'the action is past' without paying attention to the surrounding context. There is evidence in the corpus that the students are treating overt time adverbials as 'automatic tense triggers', which sometimes leads them astray, as is the case in examples (7) to (9).

(7) In a nutshell, we had the 'chance' to be far from the bombing and other raids, and so we did not pay very much attention to what happened; our life, here, has *never* been affected $was never affected$ by these events. On the contrary, the Iraqi people lived the embargo against the country (...).
(8) *For a long time,* Europeans have thought $thought$ they were more equal than any other people. They killed and used as objects millions of 'savage niggers' as they called them.
(9) *Recently,* in the area of Verviers, two drug addicts released on parole, have raped $raped$ and killed a couple of young adolescents.

The learners' behaviour is probably teaching-induced. A great many English grammars over-emphasise the role of adverbials in their presentation of tenses. This is clearly the case in Alexander's (1990:130) presentation of the two basic uses of the simple present perfect:

1. Actions beginning in the past and continuing up to the present moment:
   - with time references like *before (now), ever, never...before, up till now, so far* (...)
   - with *since/for* (...)

2. Actions which happened at an unspecified time in the past:
   - with no time reference at all (...)
   - with references to recent time, like *just, recently, already, still, yet* (...)
   - repeated/habitual actions (...).

A similar approach is taken by Swan & Houdart (1994:283): "On emploie le present perfect avec des adverbes comme *ever, never, already, yet, so far,* qui expriment l'idée de 'jusqu'à présent'". [6]

In all the above-mentioned examples, learners have clearly adopted a sentence-level approach to tenses. They have disregarded one of the major principles governing tense use, namely tense continuity, which Godfrey (1980:94-5) describes as follows: "once a tense is used representing a particular temporal reference central to the topic, the tense will continue until the topic with which it is associated is exhausted. (...) Second language speakers must control tense formation, but they must also retain and attend to the identity of tense continuities they establish if their production is to be judged acceptable."

The cohesive force of tenses is largely neglected in English grammars and ELT publications in general. There are exceptions however. Johansson & Lysvag (1986: 216) open the section on 'Tense choice' with the following statement: "Tense forms locate the action, process, or state expressed by the verb with respect to the moment of coding (speaking, writing) or some other point of reference established in the text. As the same time perspective normally characterizes longer stretches of text than the sentence, the choice of tense forms makes a significant contribution to textual cohesion". In his book on discourse for ELT teachers, Cook (1989: 15) lists verb form - and in particular verb tense form - as one of the 7 cohesive devices that make texts hang together. Similarly, Hannay & Mackenzie (1996:190) stress the importance of tense cohesion and warn learners that "as soon as a change of tense occurs, there will be a break in coherence: the reader will get the impression that a new section of the text is beginning". Their examples of unacceptable tense shifts are extracted from advanced Dutch-speaking EFL learners. This would tend to suggest that tense discontinuities are a cross-linguistic feature of advanced EFL writing. However, the problem is likely to be more acute for French-speaking learners because chronological shifts are an important feature of French tense usage. Chuquet & Paillard's (1987: 93) contrastive book contains numerous examples which highlight the heterogeneity of tenses in French texts and the more homogeneous nature of English tense usage.

5.2. L1-L2 discrepancies

While some of the difficulties encountered by advanced EFL learners seem to be shared by several national groups, others are L1-dependent. This is because, as stated by McCarthy (1991:62) "tense and aspect vary notoriously from language to language". Though this fact is widely acknowledged in the literature, it is rarely taken on board in ELT publications, which usually target all EFL learners irrespective of their mother tongue. A large number of tense errors in the learner corpus are transfer-related. In this section I will outline two particular problem areas.

5.2.1. Past Perfect

The following passages contain typical instances of past perfect misuse by French learners.

(10) Servicemen are badly paid, waste their time and forget during one year what they <u>had learned</u> $learnt$ at school.
(11) A man (or woman) who has grown up in a working-class district and who has never learned to work because his parent <u>had</u> not <u>put</u> $did not put$ him at school stands a great chance to become a criminal.
(12) In any case, when people do something for which they have no ambition, the results are often disastrous in the sense that very few remains of what <u>had been learnt</u> $was learnt$.

These errors are due to a major difference between the English and the French past perfect. Whilst the former exclusively refers to a past time earlier than another past time, the latter can also express anteriority to the present time. Generally speaking, the French past perfect has a wider range of uses than its English counterpart. Indeed, contrastive data show that only one third of the instances of the

[6] "The present perfect is used with adverbs such as *ever, never, already, yet, so far,* which express the idea of 'up to now'". (my translation)

French past perfect are translated by a past perfect in English (see van Raemdonck 1992:210).

As English grammars make it clear that the English past perfect is a past in the past, one could expect learners to infer that its use in contexts such as (10-12) is precluded. However, experience shows that learners often fail to make these inferences. A contrastive approach is necessary to make them aware of the L1-L2 discrepancy.

5.2.2. Process vs result

Many French-English contrastive books oppose the static nature of French to the more dynamic nature of English, more particularly the fact that the French tense system is more often used to express the result of a process, unlike English, which tends to make explicit the process itself (see Vinay & Darbelnet 1995, Chuquet & Paillard 1987, Van Hoof 1989). The errors illustrated in examples (13) to (15) are quite possibly a manifestation of this difference.

(13) The social rights need to be reinforced so we do not think they are acquired $have been acquired$ forever. Things can change.
(14) Some people cannot stand black people because - it's proved $has been proved$ - unconsciously, people are afraid of darkness.
(15) The social situation in which people find themselves is sometimes very hard to bear. This system has been adopted in Belgium but with a very limited scope. Only one or two districts agree $have agreed$ to use such a system.

By using a simple present, the learners focus on the resulting state, in contexts where native speakers of English would opt for a present perfect, thereby stressing the process rather than its result. Translations of French texts into English provide ample illustrations of this phenomenon (see Chuquet & Paillard 1987: 83-4). Here again, a contrastive approach is necessary to understand the students' difficulty and find ways of remedying it.

6. Implications for EFL teaching

One major pedagogical implication of our study is that tenses need to be taught at discourse level. Tenses are part of cohesion, an aspect of language which, as pointed out by Cook (1989:127) "has often been neglected in language teaching, where sentences have been created, manipulated, and assessed in isolation. (...) The results of this neglect are familiar to teachers and learners at all levels, for they affect both production and processing. In production they can result in the creation of a stretch of language in which every sentence, in isolation, is faultless, yet the overall effect is one of incoherence or inappropriateness". The inappropriate tense shifts in advanced EFL writing are one manifestation of this phenomenon. The reason why the progress rate for tense usage is so low is that students have been taught tenses at sentence level and continue to 'function' at that level even when they have reached a more advanced level of proficiency. The higher progress rate for auxiliaries may well be due to the fact that a sentence level approach is often more valid for auxiliaries than for tenses.

Another factor that has a detrimental effect on tense usage is an undue emphasis on formal markers, such as *for, since, never* in the early teaching stage. Rules such as 'if there is *for* in the sentence, use the present perfect' may seem to be useful shortcuts but, in the long term, may prove to be unhelpful, partly because they reinforce the tendency to select tenses at the clause/sentence level.

A second major implication of our study is that advanced teaching of tenses ought to be partly contrastive. Initially, tenses ought to be taught non-contrastively, as a system in its own right with its internal coherence and major underlying principles. However, especially at the more advanced proficiency levels, this 'English-only' framework should be complemented with contrastive insights. Johansson & Lysvag's grammar (1986 &1987) is a particularly good illustration of this type of approach.

Computer corpora are an excellent source of data for teaching tenses. They enable teachers to present tenses in authentic texts and allow them to manipulate them in various ways to generate contextualized exercises. Native English corpora can be used to design gap-filling exercises, where the students are requested to provide the appropriate tense for each bracketed verb in the text. Gethin (1992) and Maniez & Melter (1992) contain numerous exercises of this type. Note, however, that the authenticity

of the data is no guarantee of pedagogical soundness. M. Goodale's (1995) *Collins Cobuild Concordance Samplers. 4: Tenses* uses authentic texts from the Bank of English to present tenses. However, the concordance format chosen by the author - the KWIC format - forces students to adopt a very local approach to tense and is therefore a potential source of what James (1998:193) calls 'exercise-based induced errors'.[7] Bilingual corpora are also potentially very useful to highlight and practise L1-L2 diffferences. On the basis of bi-texts, ie aligned bilingual texts, with verb forms highlighted in the original texts and their translations deleted in the target text, students could be requested to fill in the gaps with the verb in the appropriate tense. Error-tagged learner corpora are a very rich source of exercises too. The correct forms in passages such as (1-6) above could be deleted and students asked to provide them.

7. Conclusion

Being 'special corpora' (Sinclair 1995:24), computer learner corpora quite naturally call for their own techniques of analysis. Corpus annotation is a case in point. The traditional types of annotation (part-of-speech tagging, syntactic tagging, semantic tagging) need to be supplemented with new types of annotation, such as error tagging, which are specially designed to cater for the anomalous nature of learner language. This preliminary investigation of verb tense errors in advanced learner writing demonstrates the potential of this new technique for English teaching purposes. Error tagged corpora will help ELT specialists to identify learners' difficulties at any given level of proficiency for any given learner population with a much higher degree of accuracy and thereby help them produce much more closely targeted and hence more effective pedagogical tools.

Acknowledgements

I gratefully acknowledge the support of the Belgian National Scientific Research Council (FNRS). I would also like to thank Estelle Dagneaux and Sharon Denness for their contribution to the error tagging stage of the investigation and Stephanie Petch-Tyson for insightful comments on an earlier version of this article.

Reference

Alexander L.G. 1990. *Longman English Grammar Practice for intermediate students.* Harlow: Longman.

Chuquet H. & Paillard M. 1987. *Approche linguistique des problèmes de traduction. Anglais ⇔ Français.* Paris: Ophrys.

Cook G. 1989. *Discourse.* Oxford: Oxford University Press.

Dagneaux E., Denness S., Granger S., F. Meunier. 1996. *Error Tagging Manual. Version 1.1.* Centre for English Corpus Linguistics. Louvain-la-Neuve: Université Catholique de Louvain.

Dagneaux E., Denness S. & Granger S. 1998. Computer-aided error analysis. *System* 26/2, 163-174.

Gethin H. 1996. *Grammar in Context. Proficiency Level English.* Harlow: Longman.

Godfrey D. L. 1980. A Discourse Analysis of Tense in Adult ESL Monologues. In (ed.) *Discourse Analysis in Second Language Research,* ed. by D. Larsen-Freeman. 92-110. Rowley: Newbury House.

Goodale M. 1995. *Collins Cobuild Concordance Samplers 4: Tenses.* London: HarperCollins.

Granger S. (ed.) 1998. *Learner English on Computer.* London & New York: Addison Wesley Longman.

Granger S. & Tribble C. 1998. Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning. In *Learner English on Computer* ed. by S. Granger. 199-209.

[7] For a description of concordance-based exercises exposing both their strengths and their potential dangers, see Granger & Tribble 1998).

London & New York: Addison Wesley Longman.

Hannay M. & Mackenzie J. L. 1996. *Effective writing in English . A resource guide.* Groningen: Martinus Nijhoff.

James C. 1998. *Errors in Language Learning and Use. Exploring Error Analysis.* London & New York: Longman.

Johansson S. & Lysvag P. 1986. *Understanding English Grammar. Part 1: An Overview.* Norway: Universitetsforlaget AS.

Johansson S. & Lysvag P. 1987. *Understanding English Grammar. Part 2: A Closer View.* Norway: Universitetsforlaget AS.

Kennedy G. 1998. *An Introduction to Corpus Linguistics.* London & New York: Addison Wesley Longman.

Leech G. 1998. Learner corpora: what they are and what can be done with them. In *Learner English on Computer* ed. by S. Granger. xiv-xx. London & New York: Addison Wesley Longman.

McCarthy M. 1993. *Discourse Analysis for Language Teachers.* Cambridge: Cambridge University Press.

Manier F. & Melter M. 1992. *English Grammar Alive!* Paris: Dunod.

Sinclair J. 1995. Corpus typology - a framework for classification. In *Studies in Anglistics,* ed. by Melchers G. and Warren B. 17-33. Stockholm: Almqvist and Wiksell International.

Swan M., Houdart F. 1994. *Pratique de l'anglais de A à Z.* Paris: Hatier.

Van Hoof H. 1989. *Traduire l'anglais. Théorie et pratique.* Paris & Louvain-la-Neuve: Duculot.

van Raemdonck A. 1992. *The expression of past time in French and English Journalese.* Unpublished MA Dissertation, Louvain-la-Neuve: Université Catholique de Louvain.

Vinay J.P. & Darbelnet J. 1995. *Comparative Stylistics of French and English. A methodology for translation.* Amsterdam & Philadelphia: John Benjamins.