The passive in learner English

Corpus insights and implications for pedagogical grammar

Sylviane GRANGER
Université catholique de Louvain

Abstract

The emergence of corpus-based research has revealed novel insights into the workings of the passive. In this article I give an overview of the new features that have emerged from the analysis of both native and learner corpora and check some of the findings, in particular those related to frequency of use and overpassivization, on the basis of two large learner corpus collections which contain written data from many different learner populations: the *International Corpus of Learner English (ICLE)* (Granger et al 2009) and the *International Corpus Network of Asian Learners of English (ICNALE)* (Ishikawa 2011). I also examine a range of pedagogical grammars of English to assess to what extent they have integrated corpus-based insights and make suggestions for improved coverage of the passive.

Keywords

Passive, *ICLE, ICNALE*, pedagogical grammar, overpassivization, underuse, misuse

I Introduction

The passive has always had pride of place in English grammar and linguistics. It features prominently in all English grammars, both reference and pedagogical, and its role in language has been described thoroughly in a range of linguistic theories. For a long time – and still today in some cases – descriptions of passive structures were primarily based on intuition and the focus was mainly, if not exclusively, on the structural aspect of the passive, i.e. the active-passive transformation (*the local authority built the house* / *the house was built by the local authority*), the range of passive verbal forms (from simple ones like *was built* to much rarer complex ones like *will be being interviewed* or *might have been being interviewed*) and the different types of passive structures: monotransitive (*the school was opened in 1969*), ditransitive (*they were given free bus tickets*), prepositional (*the children were looked after by a nurse*)

and impersonal (*it is generally believed that music has a healing effect*). Sections on the factors that influence the selection of the passive over the unmarked active voice and the registers the passive is associated with have tended to be much briefer. As regards second language acquisition (SLA), studies of the passive have primarily relied on elicitation data, often in the form of grammaticality judgment tests, and hence contain little information on authentic language use. This is changing with the emergence of corpus linguistics which, for the first time, provides the data and tools to give much more precise insights into passive use. This article pursues three objectives: (1) present the picture of the passive that has emerged from the analysis of native and learner corpora; (2) revisit some of the previous findings on the basis of two large learner corpus collections; and (3) examine to what extent these new insights have been heeded in recent pedagogical grammars and make suggestions for improved coverage.

II Corpus-based studies of the passive

2.1 Native corpora

The publication that marked a turning-point in studies of the passive voice is Svartvik's (1966) book length study based on a corpus covering eight text types, seven of which represent written registers (science, news, arts, sports, novels, advertising, plays). Inspired by Svartvik's work I conducted a similar study (Granger 1983) based on five spoken registers (conversation, discussion, interview, commentary and oration). These two corpus-based studies, together with the many others that have followed, have led to a much better understanding of the passive, particularly in connection with the following issues:

- Frequency of the different types of passive (*be*-passive vs. *get*-passive; monotransitive, ditransitive, prepositional, impersonal structures; agentful vs. agentless; etc.). Two important findings are that 80-90% of the passives are agentless and that *get*-passives are rare.
- Register associations of the passive: text types can be ranked along a cline from the highest passive rate (scientific writing) to the lowest (informal conversation). In Biber's (1988) multi-dimensional model, the passive is associated with Dimension 5 which distinguishes texts with an abstract vs. non-abstract information focus. A large number of corpus-based studies have analyzed the major role played by the passive in academic writing and have taken issue with the 'avoid the passive' prescription that is still found in many usage guides.
- Lexical associations of the passive: there are strong associations between active-passive constructions and lexical choices (cf. Biber et al 1999; Gries & Stefanowitsch 2004). Some verbs display strong passive attraction, while others are characterized by passive repulsion. Table 1 provides a sample of 12 verbs that

illustrate the extended range of passive ratios in the academic section of the *BNC Baby*[1] (from 0.8% for *want* to 68.2% for *oblige* through the verbs highlighted in grey that display the standard ratio in academic writing, viz. c. 20%).[2] More fine-grained analyses show that the various meanings of polysemous verbs have different passive ratios. Haas (2008: 274) observes that the five meanings of the verb *meet* have the following passive ratios: 0% ('come together'), 2% ('come across'), 21% ('standard'), 72% ('arrival') and 81% ('reaction').

Table 1: Passive ratio in the academic section of the *BNC Baby*

lemma	npassive	nlemma	passive ratio
oblige	15	22	68.2%
deem	9	15	60%
entitle	31	56	55.3%
expect	143	267	53.5%
give	311	1368	22.7%
make	380	1763	21.5%
reach	32	157	20.4%
show	224	1179	19%
learn	7	165	4.2%
receive	8	240	3.3%
attend	2	73	2.7%
want	2	250	0.8%

2.2 Learner corpora

The passive has been the subject of a large number of corpus-based studies of learner language. Two main findings have emerged from these studies. First, most learner populations significantly underuse the passive. For example, Hinkel (2004) finds consistent underuse across the six learner groups he analyses (Chinese, Japanese, Korean, Indonesian, Vietnamese and Arabic). Similar findings are reported in many studies, among which Granger (1997) and Xiao (2007). As pointed out by Gilquin (2008: 6), the underuse of the passive seems to be a universal feature of interlanguage, which may be related to learners' preference for the unmarked rather than marked option or a personal rather than impersonal style. However, this tendency may be reinforced when the corresponding structure in the learners' mother tongue is much less frequent than in English. Second, many learner populations prove to misuse the passive, one of the main types of error reported in the literature being overpassivization of verbs such as *occur, disappear* or *suffer* (Oshita 2000, Cowan et al 2003).

III The passive in *ICLE* and *ICNALE*

Because many previous corpus-based studies of the passive are based on small samples and a limited number of subjects, it is useful to revisit some of the results on the basis of large 'multi-L1 learner corpora' (Granger 2012: 12.), i.e. corpora which cover learners from several mother tongue backgrounds. The two largest collections of this type are the *International Corpus of Learner English. Version 2 (ICLEv2)* (Granger et al 2009) and the *International Corpus Network of Asian Learners of English (ICNALE)* (Ishikawa 2011) (see Table 2 for details of the two corpora). The two corpora are to some extent complementary as they cover partly different language backgrounds: mostly European for *ICLE* and exclusively Asian for *ICNALE*. They also differ in several other respects such as average text length and task conditions. One difference that will play a part in our analysis is that the *ICLE* subjects wrote essays on a large number of topics while only two topics are used in *ICNALE*: 'Smoking should be completely banned at all the restaurants in the country' and 'It is important for college students to have a part time job'. Another important difference is that *ICLE* only includes data from learners of English as a Foreign Language while *ICNALE* also contains data from learners of English as a Second Language. In this section I will compare insights gained from the two corpora in terms of (1) frequency of passives and (2) overpassivization errors.

Table 2: Details of *ICLEv2* and *ICNALE*

	ICLEv2	**ICNALE**
Learners' mother tongue backgrounds (ICLE) and countries of origin (ICNALE)	Bulgarian, Chinese[3], Czech, Dutch, Finnish, French, German, Italian, Japanese, Norwegian, Polish, Russian, Spanish, Swedish, Tswana, Turkish	China, Indonesia, Korea, Japan, Hong Kong, Pakistan, The Philippines, Singapore, Taiwan, Thailand
Total number of words	3.7 million words	1.2 million words[4]

3.1 Frequency of the passive in *ICLE* and *ICNALE*

Using the search interface integrated in the CD-ROM I extracted all the occurrences of <Vbe> <VVN> (hereafter referred to as *be Ved*), i.e. all the occurrences of the lemma BE followed by a past participle, from *ICLEv2* (Granger et al 2009). I applied the same method to two native corpora: the academic section of *BNC Baby*, a corpus of expert writing, and the *Louvain Corpus of Native English Essays (LOCNESS),* a corpus of novice argumentative writing (American undergraduate students). One disadvantage of the method is that it will not retrieve passive sequences interrupted by an adverb or some other element (e.g. *it has been recently noted*). Though imperfect the method

provides a good starting-point for more in-depth analyses. The results are shown in Table 3. As compared to the expert native corpus (*BNC Baby*), all learner populations underuse the passive but the degree of underuse varies greatly across the learner populations. It should be noted that the *LOCNESS* writers also significantly underuse the passive, a feature which is probably due to the fact that they are novice writers (although one cannot exclude the impact of the stigma attached to the passive in freshman composition classes).

Table 3: Frequency of *be Ved* forms in *ICLEv2, BNC Baby_ACAD* and *LOCNESS* (freq./100,000 words)

Corpus	Relative frequency (100,000 words)
BABY BNC_ACAD	**1436**
Tswana	1257
Dutch	1215
LOCNESS	**1081**
Turkish	1050
Finnish	1005
Spanish	998
Polish	987
French	986
Chinese	892
Russian	861
Swedish	843
Italian	821
Bulgarian	816
German	810
Czech	796
Norwegian	787
Japanese	782

Underuse of the passive is also noticeable in *ICNALE* (cf. Table 4). It is interesting to observe that all the varieties – both outer circle and expanding circle – underuse the passive compared to the inner circle varieties (*BNC* and *LOCNESS*). However, the outer circle varieties (Hong Kong, Pakistan, Philippines and Singapore) cluster at the top: they are characterized by a higher frequency of the passive than the expanding circle varieties. As regards the latter, a comparison with Table 3 shows that the underuse is more marked in *ICNALE* than in *ICLE*. This may be due to several factors, among which different task conditions (untimed for 62% of the *ICLE* subjects vs. 100% for

ICNALE) and differences in proficiency level (mainly B1 level of the Common European Framework of Reference for Languages in *ICNALE* vs. B2-C1 for *ICLE*).

Table 4: Frequency of *be Ved* forms in *ICNALE, BNC Baby_ACAD* and *LOCNESS* (freq./100,000 words)

Corpus	Relative frequency (100,000 words)
BABY-BNC-ACAD	**1436**
LOCNESS-US-arg	**1081**
The Philippines	942
Singapore	940
Hong Kong	912
Pakistan	873
China	721
Korea	710
Japanese	615
Indonesia	574

ICNALE is subdivided into proficiency bands, which makes it possible to link up frequency of passive use and proficiency level. As the passive is quite a complex structure and the corresponding structures in the learners' L1s are often less frequent than the passive in English, we can expect the more advanced learners in *ICNALE* to use the passive more than the less advanced ones. To check this hypothesis I analysed the Japanese component of *ICNALE* (*ICNALE-JP*). As shown in Table 5, the results do not support the hypothesis; the frequency even seems to decrease as proficiency increases. However, a close look at the data shows that these figures are misleading. First, as shown in Figure 1, it is consistently the essays dealing with smoking (SMK) that are characterized by higher passive frequency, irrespective of the proficiency level. This is mainly due to the repetitive use of *be banned*, which itself is borrowed from the essay prompt 'Smoking should be completely banned at all the restaurants in the country'. Forms of *be banned* account for the following proportions of the SMK essays: 20% at level A2, 28% at B1.1, 37% at B1.2 and up to 47% at level B2. It is important to note that the other prompt ('It is important for college students to have a part time job' – referred to as PTJ) does not contain any passive form. This finding shows that although constraining topics in learner corpus collection can be an asset for some types of research (e.g. to measure vocabulary richness), it may run the risk of skewing the results for the analysis of heavily lexicalized aspects of grammar such as the passive (and may also affect other features such as the number and types of lexical bundles, cf. Ishikawa 2009). Second, a close analysis of the PTJ concordance lines reveals the

existence of a large number of erroneous *be Ved* sequences. One frequent type of error, especially at the lower levels, is the overextension of the passive form to active contexts (cf. examples 1-3). The slight downward trend noticeable in Figure 1, both for SMK and JPT, may therefore be partly due to an increasing mastery of verb morphology.

(1) Secondly, a part time job *can be improved* student's communication skill (PTJ_A2)
(2) We may *be noticed* how important money is and we may appreciate my parents who work hard for family every day (PTJ_B1.1)
(3) Then I was playing with my friends every day and night. So, I *was lost* much money (PTJ_B1.1).

Table 5: Frequency of *be Ved* forms in *ICNALE_JP* (freq./100,000 words)

ICNALE_JP	Relative frequency (100,000 words)
Total A2	653
Total B1.1	608
Total B1.2	548
Total B2[5]	523

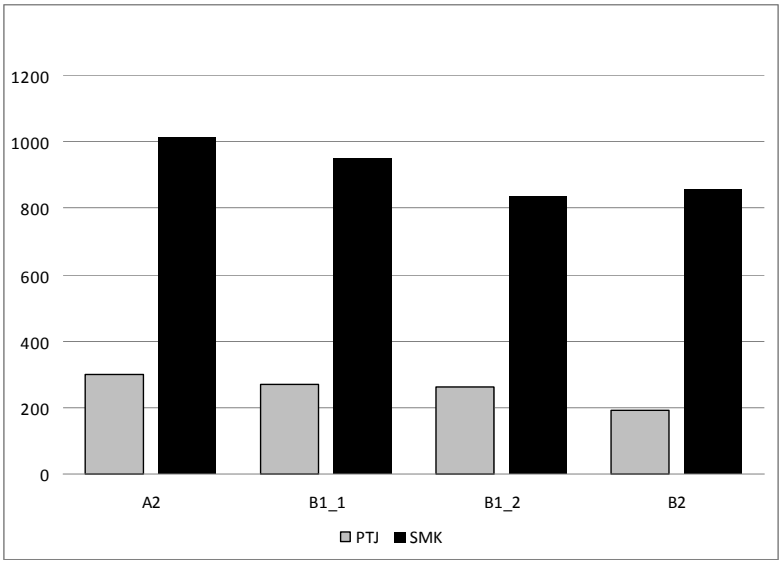

Figure 1: *ICNALE_JP*: breakdown of *be Ved* forms per proficiency level and topic (freq./100,000 words)

3.2 Overpassivization errors in *ICLE* and *ICNALE*

With a view to investigating overpassivization errors in *ICLE* and *ICNALE*, I extracted all instances of the lemma BE immediately followed by the past participle form of the following ten verbs: *arrive, die, occur, become, happen, disappear, seem, appear, exist,*

suffer. Based on the literature I expected a higher number of these errors in Asian learner writing than in the other learner populations as overpassivization has been reported as a major source of difficulty for several Asian learner populations, especially Chinese, Korean and Japanese (Yip 1995, Oshita 2000 and Cowan et al 2003). As shown in Tables 6 and 7, this hypothesis is only partly confirmed. The Chinese learners and Japanese learners in *ICLE* and the Korean and Hong Kong learners in *ICNALE* are the populations with the highest number of overpassivization errors. However, many learner populations, including Asian ones, have very few errors with the selected verbs and several do use them erroneously at all (7 learner populations in *ICLE* and 1 in *ICNALE*). Further research is needed to link up the phenomenon of overpassivization and learners' L1 background and proficiency level. In future studies it is advisable not to isolate overpassivization errors from other verb morphology errors. The overpassivization errors illustrated in (4) and (5) should be investigated alongside the 'underpassivization' errors (i.e. lack of passive morpheme) illustrated in (6) and (7).

(4) Many no smokers *are existed* in the world (ICNALE_KOR)
(5) about 600, 000 people *are died* by secondhand smoking (ICNALE_KOR)
(6) human beings *are exist* with their own rights (ICLE_TR)
(7) Some *are die* from traffic accident (ICLE_CH)

Table 6: Overpassivization errors in ICLE (freq./100,000 words)

CH	JA	TU	IT	TS	SP	CZ	RU	FR	BU	DU	FI	GE	NO	PO	SW
11.8	9	8.5	3.5	3	1.5	1.4	1.2	0.4	0	0	0	0	0	0	0

Table 7: Overpassivization errors in ICNALE (freq./100,000 words)

KOR	HKG	IDN	PAK	THA	JPN	CHN	TWN	PHL	SIN
16	14.1	9.7	6.4	4.5	3.9	3	2.2	2	0

IV Implications for pedagogical grammars

All ELT grammars devote a whole chapter to the passive. With a view to assessing to what extent they integrate the insights generated by corpus-based studies, I carried out a detailed analysis of the passive sections in 11 recent pedagogical grammars for higher intermediate/advanced learners of English (cf. list in Appendix 1). This turned out to be a very instructive exercise. For reasons of space only the major findings will be reported here. The main observation that can be made is that corpus-based studies of the passive have had relatively little impact on pedagogical grammars. As regards frequency, the dominance of the agentless passive is usually mentioned (although this information tends to be backgrounded) but the *get*-passive tends to be presented as

very frequent and interchangeable with the *be*-passive. Register preferences are either totally absent or given very cursory treatment. One notable exception is Celce-Murcia & Larsen-Freeman (1999) who give a good description of the distribution of the passive across genres. Lexical aspects are nearly always presented negatively as lists of verbs that cannot passivize. It is striking that the verbs that are nearly always mentioned are transitive verbs such as *have, lack, equal, fit* or *resemble* which cause less difficulty to learners than intransitive and linking verbs like *happen, occur* or *become* which are rarely mentioned. For example, Swan's (2005: 386) warning against the erroneous passive forms *A nice house is had by them, I'm fitted by my shoes* and *Tact is lacked by your mother* is not very useful as these verbs are very rarely passivized by learners. The only grammars that give due attention to the most error-prone verbs are Celce-Mulcia & Larsen-Freeman (1999) and Cowan (2008). The negative slant of the sections devoted to the passive is reinforced by rules that prescribe the use of the passive with some categories of verbs. For example, Carter et al (2000: 257) give the following rule: "A passive cannot be formed using stative verbs which refer to states not actions". This is a very misleading rule as many stative verbs are often used in the passive (e.g. the verb *to know* and *to believe* which are given as examples of stative verbs in the glossary of the grammar). With all these warnings against using the passive, it is no wonder that learners underuse it!

V Conclusion

Two main findings emerge from our study. First, the overwhelming majority of learners of English as a second or foreign language underuse the passive. Second, different learner populations produce different types of errors. This is clearly the case for overpassivization errors, probably due to different morphological properties of the learners' mother tongues. These two findings have major implications for pedagogical grammars. First, sections on the passive should have a much more positive slant, with examples of prototypical uses of the passive and lists of verbs that often passivize, rather than the 'can't do' approach that is still dominant in many current grammars. Second, while it is useful to include warnings in pedagogical grammars, care must be taken to include warnings that are potentially useful to a large number of learner populations and leave out warnings against errors that are L1-specific (or make explicit reference to the targeted L1 groups, cf. Cowan 2008: 410-412). Many different factors – in particular, proficiency level, task and mother tongue background – play a part in the use of the passive. A preliminary analysis such as the one reported here cannot hope to do justice to the complex interplay of these factors and the relative importance they have in the different learner populations. Future research in the field should make use

of more flexible search strings which allow for the extraction of non-contiguous *be Ved* forms. More crucially, the automatic extraction stage must be followed by careful examination of the extracted sequences in context so as to distinguish between correct passive forms and other uses of *be Ved* and thereby gain a deeper understanding of the acquisition processes at play. At this stage, however, one thing seems clear: both for theoretical and applied investigations of learner language, multi-L1 corpora like *ICLE* and *ICNALE* are invaluable resources which should not be neglected in future interlanguage research.

Notes

[1] http://www.natcorp.ox.ac.uk/corpus/babyinfo.html.

[2] Contiguous sequences of the lemma *be* followed by a past participle were extracted from the POS-tagged version of the corpus.

[3] Mainly from Hong Kong (c. 80%).

[4] This total does not include the inner circle control corpora.

[5] It is not possible to draw reliable conclusions for this level as the data size is very small (8,000 words in total).

References

Biber, D. (1988). *Variation across Speech and Writing.* Cambridge University Press: Cambridge.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan. E. (1999). *Longman Grammar of Spoken and Written English*. Harlow, England: Pearson Education.

Cowan, R., Choi, H. E and Kim, D. H. (2003), Four questions for error diagnosis and correction. *CALICO Journal* 20(3), 451-463.

Gilquin, G. (2008). Combining contrastive and interlanguage analysis to apprehend transfer: detection, explanation, evaluation. In G. Gilquin, S. Papp & M. B. Diez-Bedmar (eds.), *Linking up Contrastive and Learner Corpus Research* (pp. 3-33). Amsterdam & Atlanta: Rodopi.

Granger, S. (1983). *The be + past participle construction in spoken English with special emphasis on the passive.* Amsterdam: Elsevier Science Publishers.

Granger S. (1997). Automated retrieval of passives from native and learner corpora: precision and recall. *Journal of English Linguistics* 25(4), 365-374.

Granger, S. (2012). How to use foreign and second language corpora. In A. Mackey & S.M. Gass (eds.) *Research Methods in Second Language Acquisition* (pp. 7-29).

Granger, S., Dagneaux, E., Meunier, F., Paquot, M. (2009). *The International Corpus of Learner English. Handbook and CD-ROM. Version 2.* Louvain-la-Neuve: Presses universitaires de Louvain.

Gries, S. Th. & Stefanowitsch, A. (2004). Extending collostructional analysis. A

corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1), 97-129.

Haas, F. (2008). Symmetric verbs and constraints on passivization: An English-German comparison. *Zeitschrift für Anglistik und Amerikanistik* 56(3), 269-282.

Hinkel, E. (2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research* 8(1), 5-29.

Ishikawa, S. (2009). Phraseology overused and underused by Japanese learners of English: A contrastive interlanguage analysis. In K. Yagi & T. Kanzaki (eds.). *Phraseology, Corpus Linguistics and Lexicography: Papers from Phraseology 2009 in Japan* (pp.87-100). Nishinomiya, Japan: Kwansei Gakuin University Press.

Ishikawa, S. (2011). A new horizon in learner corpus studies: The aim of the ICNALE project. In G. Weir, S. Ishikawa, & K. Poonpon (eds.), *Corpora and language technologies in teaching, learning and research* (pp.3-11). Glasgow, UK: University of Strathclyde Publishing.

Oshita, H. (2000). What is happened may not be what appears to be happening: a corpus study of 'passive' unaccusatives in L2 English. *Second Language Research* 16(4): 293-324.

Svartvik, J. (1966). *On Voice in the English Verb.* The Hague: Mouton.

Yip, V. (1995). *Interlanguage and Learnability: From Chinese to English*. Amsterdam: Benjamins.

Appendix 1: List of ELT grammar books

Azar, B.S. (1999). *Understanding and using English Grammar*. New York: Longman.

Carter, R., Hughes, R. & McCarthy, M. (2000). *Exploring Grammar in Context*. *Grammar reference and practice upper-intermediate and advanced.* Cambridge: Cambridge University Press.

Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The Grammar Book. An ESL/EFL Teacher's Course.* Boston: Heinle & Heinle.

Cowan, R. (2008). *The Teacher's Grammar of English. A coursebook and reference guide.* New York: Cambridge University Press.

Eastwood, J. (2006). *Oxford Practice Grammar*. Oxford: Oxford University Press.

Foley, M. & Hall, D. (2003). *Longman Advanced Learners' Grammar. A self-study reference & practice book with answers.* Harlow: Longman**.**

Hewings, M. (2005). *Advanced Grammar in Use. A self-study reference and practice book for advanced students of English.* Second edition. Cambridge: Cambridge University Press.

Lester, M. (2008). *McGraw-Hill's Essential ESL Grammar. A Handbook for Intermediate and Advanced ESL Students*. New York: McGraw-Hill.

Penston, T. (2005). *A Concise Grammar for English Language Teachers*. Greystones: TP Publications**.**

Swan, M. (2005). *Practical English Usage*. Oxford: Oxford University Press**.**

Vince, M. (2008). *Macmillan English Grammar in Context. Advanced*. Oxford: Macmillan.

Please refer to this article as follows:

Granger, S. (2013). The passive in learner English: Corpus insights and implications for pedagogical grammar. In S. Ishikawa (ed.) *Learner Corpus Studies in Asia and the World. Vol. 1. Papers from LCSAW2013*. Kobe: School of Languages and Communication, Kobe University, 5-15.