Semi-specialized corpora of written French as a resource in language teaching and learning

Angela Chambers and Victoria Kelly
University of Limerick

Abstract

This study is based on data provided by a corpus of 1,144,668 words of contemporary written standard French of a semi-specialized nature, consisting of journalistic texts relating to the introduction of the Euro. The aim is to investigate the potential of a corpus such as this as a resource for language teachers and learners in an area such as French for business, where competence in the expression of quantitative data may be required. Through an analysis of the verbs and prepositions used in conjunction with the percentage symbol, the aim is twofold: firstly to show how the information obtained from the corpus can complement the contribution of the grammar and the coursebook, providing a resource with a variety of examples of actual use; and secondly to raise the question of how information such as this can be made available to teachers and learners in an environment which fosters inductive learning and learner autonomy.

Introduction

Despite the existence of an increasing amount of research in corpora and language learning and teaching, it remains the case that most developments are based on the English language, with both written and spoken corpora available to the researcher. In other languages this is not yet the case, with the result that in French, for example, written corpora are still largely literary in nature, with no equivalent to the vast English-language reference corpora. Moreover, as the most recently published research in corpus linguistics in French tends to focus on spoken corpora of contemporary language use (see, for example, Blanche-Benveniste 2001a, b), there is still a dearth of research on the use of written corpora of standard, contemporary, non-literary French in the teaching and learning context. This is not to say that researchers in the French language are

unaware of the potential in this area, as Habert's comment on the potential of texts in electronic form such as *Le Monde* on CD-ROM amply demonstrates:

> il devient alors possible d'étudier l'évolution de certains phénomènes langagiers: néologismes, emplois privilégiés à un moment donné de certains suffixes ou préfixes, etc., un peu comme les éditions papier de certains dictionnaires d'usage ... servent de <<sonde>> sur le lexique et ses changements. Les CD-ROMs du journal *Le Monde* permettent aujourd'hui de telles analyses pour le français. (Habert 2000: 13)

This study aims to contribute to research in this area, by examining how a corpus of approximately one million words of written French of a semi-specialized nature (journalistic texts relating to the introduction of the Euro) compares as a language learning and teaching resource to its more traditional counterparts, such as the textbook or grammar.

As with the majority of publications in corpora and language learning and teaching in recent decades, the pedagogical context of this study is greatly influenced by constructivist theories of learning in general, and by the communicative approach to language learning and developments in the area of learner autonomy in particular. The ambition to provide learners with resources to enable them to base their communication on real language use, and to make decisions concerning their own learning, led to a re-appraisal of the traditional printed coursebook (Tomlinson 1998) and the grammar book as resources, with their basis in native speaker intuition and their invented examples of use. While the inclusion of authentic, contemporary, non-literary texts in coursebooks represented a step forward, it still did not allow for a significant level of learner autonomy. Moreover, the learning of grammar, lexis, and usage was of necessity based on the content of the coursebook or grammar, with very limited potential for inductive learning based on a rich variety of examples of real language use. The availability of corpora has radically altered this situation, particularly if both teachers and learners have access to them. However, the research necessary to underpin such access is still in its infancy. While recent empirical studies on the results of providing direct access to corpora by learners certainly reveal its benefits (see, for example, Cobb 1997 and

Kennedy and Micelli 2001), it is also important to note that, even in a relatively small corpus, the number of occurrences of a given item may be daunting for a learner who has no specific intention to acquire expertise in corpus linguistics. The way forward may lie also in other directions, such as online 'worksheets', where the learner can inductively choose a path to follow through classifications of examples provided by the researcher/materials developer. Whatever solutions emerge, studies of language use in both general and more specialized contexts are necessary to provide the data for such developments. As we shall see, the results of this study would easily lend themselves to such a development.

The small, semi-specialized corpus

While large reference corpora are undoubtedly the most reliable source of information on language use for the researcher and teacher, they also present problems not only for the learner but also for the teacher who may be interested in using corpora as a resource, but unable or unwilling to devote a large amount of time to the necessary training. As Aston (1997) has pointed out, using a small corpus has many advantages for those without the training to enable them to work with large corpora, even if access to the larger collections of texts were easily available. While Aston defines the small corpus as typically between 20,000 and 200,000 words, and large corpora as starting at 1,000,000 (Aston 1997: 54), we argue here that a corpus of 1,000,000 words can also be included in the category of small corpora, and is likely to provide a richer source of examples of a given item than its smaller counterpart, which can at times prove to be of no help at all to a learner for certain items with very few occurrences. Even in the area of general language, it has been pointed out that 'huge corpora are not necessary for language-teaching purposes. A modest corpus of a million or so words is certainly enough to make a valuable teaching aid and is realistically within the reach of most teaching institutions' (Dodd 1997: 131). Aston notes the following advantages of small corpora over their much larger counterparts: they are easier to manage; easier to become familiar with; easier to interpret; easier to construct; easier to reconstruct; more clearly patterned; and their limits are clearer. He concludes:

> To sum up, I would suggest that work with small spe-
> cialized corpora can be not only a valuable activity in its
> own right, as a means of discovering the characteristics of
> a particular area of language use, but also an instrument to
> help and train learners to use larger ones appropriately.
> (Aston 1997: 61)

In addition, the specialized or semi-specialized corpus will
provide examples relevant to the context of the learner's interest, a
particular advantage in further and higher education, where the lan-
guage learning is often intended to be relevant to a particular disci-
pline, such as business studies, engineering, or tourism.

The level of specialization necessary for language learning
materials was the subject of a considerable amount of research in
the 1980s, with semi-specialized texts receiving particular attention
(see, for example, Hutchinson and Waters 1988). In much teaching
of French for business, for example in the first and second years of
a degree programme, neither the language teacher nor the learner is
a specialist in economics, marketing, or any discipline within the
area of Business Studies. This type of language class is a long way
from the accepted definition of Languages for Specific Purposes
(LSP) as specialist-to-specialist communication (Sager et al. 1980:
68). It is closer to popularization level, the first step on a four-level
scale of specialization as defined by Balboni (1986: 6), or to an
early point on the continuum from popularization to specialization
in Varantola's schema (1986: 11). To define their situation in the
research context of genre analysis (Swales 1990), both teacher and
learner belong to the discourse community of language learning for
business, not that of researchers or practitioners in the world of
business. In this context, as Hutchinson and Waters (1988: 181)
have shown in the scientific context, semi-technical rather than very
technical or general texts, can provide valuable teaching and learn-
ing resources. It is for these reasons that this study concentrates on
the small, semi-specialized corpus as a resource. The term semi-
specialized is used here to cover LSP texts which are made avail-
able to the educated general public, as in a newspaper such as *Le
Monde*. Because its sources include a variety of types of journalistic
texts relating to the Euro, this corpus includes both texts at popu-
larization level and others requiring familiarity with the discourse
of economics. This article thus includes examples at popularization
level and also more specialized concordance results. However, the

corpus is of a sufficient size to enable the materials designer or teacher to produce exercises based on results at the level appropriate to the learners.

This study is based on data provided by a corpus of 1,144,668 words of contemporary written French of a semi-specialized nature compiled by Victoria Kelly. The journalistic texts which make up the corpus are taken from three consecutive years of the French daily *Le Monde* (1999, 2000, 2001) beginning with the year 1999 which saw the launch of the Euro in the world's currency markets and ending on January 1, 2002, when the Euro entered general circulation and became legal tender for 12 of the EU member states. These three sub-corpora, which make up the one million-word corpus, consist of approximately 381,000 words each. The criteria imposed for the selection of suitable articles were that the articles had to be published in one of the three chosen years and had to deal with the topic of the Euro in any section of the newspaper. The concordancing package used to analyse the data is *Wordsmith Tools* (Scott 1999).

Conveying monetary information in written French

Learners of French at undergraduate level, even after studying the language throughout secondary school, often experience difficulties with expressions relating to numbers in general, and percentages in particular. For example, native speakers of English tend to omit the preposition *de* in a phrase such as *Le taux d'inflation est de 3%* 'The inflation rate is 3%'. In addition, phrases relating to changes in percentages also pose problems, producing either grammatical errors or the use of words or phrases which a native speaker would not use, such as *plus haut* 'higher' and *plus bas* 'lower' instead of *supérieur* and *inférieur*. There is thus a need for teachers and learners to have access to examples of actual discourse in which they can identify lexical patterns used by native speakers. In this study we have chosen the percentage symbol, %, as the search item, in order to identify which lexical items and which lexical patterns can be observed when conveying information relating to percentages, a skill frequently required in the context of business. The resulting data (6,710 occurrences) could be analysed from a variety of perspectives, for example in relation to prepositions, verbs, adjectives, conjunctions, etc. occuring frequently with the symbol. In this study

we have decided to concentrate on verbs and prepositions, as the collocations and colligations can conveniently be studied together, and represent an area where learners have problems that cannot be easily rectified by consulting the grammar, dictionary, or course-book. Although definitions of 'collocation' focus on 'the occurrence of two or more words within a short space of each other in a text' (Sinclair 1991: 170), we have decided to use the term in a wider sense here, to refer to the words which occur with the percentage symbol. Partington traces the definition of the term back to Firth (1957), noting that he used it in a wider sense 'to refer to the co-occurrence of items at all grammatical levels, not just the word level' (Partington 1998: 16) and defining it as 'phenomena of pat-terning in text' (1998: 17). While he does not explicitly extend his definition beyond lexical items, it is arguable that his phrase could justify the extension of collocation to include symbols, particularly in the case of specialized or semi-specialized corpora. This ex-tended definition is also used here in relation to colligation, by which we mean the 'tendency of an item to occur in particular grammatical company' (Aston 2001: 14).

To compare the corpus-based results with the information available in existing published resources, a number of university-level textbooks and grammars available to advanced learners of French were selected and consulted as to their presentation of the verbs and prepositions used in the expression of percentages. Ex-cept for one well-known grammar, which is now in its thirteenth edition, (Grevisse 1997: 652), the expression of percentages re-ceives little or no attention in the seven textbooks and grammars consulted, including three textbooks published in France, three in Great Britain, and one in Germany between 1990 and 2002. Fol-lowing Lawson's example (2001: 184-185), it has been decided not to cite those studies, as they cannot be criticized for this 'omission'. This is because discourse-based grammars and course-books can only become the norm when at least one large reference corpus is available, which is not the case in French. As we shall see, however, a relatively small corpus can provide a rich variety of contemporary examples of use that could form a valuable basis for a teaching and learning resource in this area to complement the grammar or course book.

An initial search carried out on the 1 million-word corpus using the character representing a percentage symbol (%) as the 'search word' resulted in 6,710 occurrences. Four categories were identified

in relation to the colligational affinity of prepositions with the percentage symbol. In Table 1, the category Verb + no preposition is followed by two prepositions encountered in order of frequency, *de* and *à*. The preposition *contre* (299 occurrences) is not included in this study, as no significant colligations with verbs were encountered.

Table 1: Overall totals per category

Category	Total number of occurrences with %
Verb + no preposition	450
Verb + preposition *de*	756
Verb + preposition *à*	526

Verb + no preposition category

This category initially reveals some obvious patterns, in that a limited number of verbs frequently collocate with the percentage symbol, the most frequently occurring being *atteindre, gagner, perdre,* and *représenter*. Examples include: *L'inflation a atteint 2,4% en juillet dans l'Euroland* 'Inflation reached 2.4% in the Eurozone in July' and *La Bourse de Paris a de nouveau cédé 1,64%* 'Paris (the French stock exchange) fell by a further 1.64%'. The relevant verbs and their frequencies are given in Table 2 below. The translations given in this and subsequent tables refer to the sense in which the terms are used in the corpus when collocating with the percentage symbol.

Further analysis of individual verbs in this category, however, reveals strong collocational preferences for the expression of the percentage. For example, when the verb *atteindre* 'to reach' is used to express the percentage, it is most frequently used when referring to growth (17 occurrences) or inflation (9 occurrences). It is most frequently used to express a probability, that inflation or growth should or could reach a certain percentage, and collocates in these instances with one of two modals, *devoir* and *pouvoir* as in the following examples: *Le taux d'inflation de la zone Euro **devrait atteindre** 2,2% cette année au lieu de 1,8% comme prévu en avril* 'The inflation rate in the Eurozone should reach 2.2% this year and not 1.8% as was forecast in April' and *L'indice des prix à la consommation **pourrait même atteindre** 3,6% en rythme annuel*

'The consumer price index could even reach an annual rate of 3.6%'. It is also important to note that the verb *atteindre,* when used to express a percentage, is often used in the past tense. In this tense, *atteindre* is used to refer to a far broader range of economic indices from the unemployment rate to the CAC index to the Euro: *Le taux de chômage **a atteint** en mars 8,4% de la population active* 'The unemployment rate reached 8.4% in March'. A further example of an individual verb in this category revealing an even more specific collocational preference is *céder* (to drop). While there are only 28 occurrences of this verb when used to express a percentage, all 28 are used with reference to fluctuations on the stock market. Examples include: *A Wall Street, l'indice Standard and Poor's 500 a cédé 8,5% en trois semaines* 'In New York, the Standard and Poor's 500 index dropped 8.5% in three weeks'.

Table 2: Word frequencies for the ten most frequent verbs occurring with the percentage symbol (no preposition)

Rank	Verb	Frequency in corpus	Frequency with %
1	*Atteindre* 'to reach'	604	101
2	*Perdre* 'to drop'	399	78
3	*Gagner* 'to gain'	265	55
4	*Représenter* 'to account for'	354	37
5	*Céder* 'to drop'	109	28
6	*Abandonner* 'to drop'	108	17
7	*Dépasser* 'to exceed'	218	13
8	*Détenir* 'to hold'	95	7
9	*Réaliser* 'to make/realize'	346	6
10	*Avoisiner* 'to be about'	14	5

Verb + preposition de *category*

The analysis of this second category (756 occurrences) reveals firstly that *de* is also used with a limited number of verbs when used to express a percentage, the most frequently occurring being *progresser, augmenter, être,* and *passer*. Examples include: *Le PIB de la zone Euro **a progressé de** 0,7% au premier trimestre* 'There was an increase of 0.7% in the GDP of the Eurozone in the first quarter' and *Les taux d'intérêt à long terme **sont passés de** 4% à*

près de 5% 'Long term interest rates have gone up from 4% to almost 5%'.

Table 3: Word frequencies for the ten most frequent verbs occurring with the percentage symbol (preposition de)

Rank	Verb	Frequency in corpus	Frequency with %
1	*Progresser* 'to increase'	368	147
2	*Augmenter* 'to increase'	493	115
3	*Être* 'to be'	4,266	109
4	*Passer* 'to move'	652	61
5	*Reculer* 'to decrease'	178	55
6	*Baisser* 'to drop'	306	34
7	*Bondir* 'to jump'	39	21
8	*Chuter* 'to drop'	102	19
9	*Remonter* 'to go up'	281	15
10	*Porter* 'to move'	330	13

It is interesting to note from the teacher's and the learner's point of view that *progresser* (147 occurrences) is more frequently used to signal a rise or increase in financial terms than *augmenter* (115 occurrences), both having the English equivalent of 'to increase'. This preference is also evident in the noun forms, with 46 occurrences of *progression* and only 24 of *augmentation*. Evidence from the corpus offers examples such as: *Le FMI s'attend à **une progression de 2% en 1999*** 'The IMF expects an increase of 2% in 1999'.

A further useful discovery for learners concerns the combination: Verb + *de* + % + *à* + %, shown in Table 4 below. In this combination the limited number of verbs encountered includes *passer*, *remonter*, *porter*, *baisser*, and *revenir*, combined with both prepositions *de* and *à* and the percentage symbol to signal an increase or a decrease in a rate. One example of this combination is: *Le taux de la facilité de dépôt **a passé de 2% à 1,5%*** 'The deposit facility rate went from 2% to 1.5%'.

Concerning the plural form of the preposition *de*, the corpus provides examples to teacher and learner of how *des* is used differently depending on whether it is placed before or after the pecentage symbol. When *des* is used after the percentage symbol (371 occurrences), examples from the corpus indicate that the press is

simply providing information, for example, ***38% des personnes interrogées*** *expriment un avis positif* '38% of those questioned express a positive opinion' and *D'après un sondage,* ***91% des patrons de petites et moyennes enterprises*** *ont avancé dans leur préparation à l'Euro* 'According to a survey, 91% of heads of small businesses are making preparations for the Euro changeover'. The advanced learner will discover, however, that when *des* is used before the percentage (98 occurrences), it appears primarily to be used to express an upper or lower limit. Furthermore, in almost all examples, the percentage being expressed refers to the level of inflation. This expression appears to be restricted to the verbs *franchir* 'to pass' (12 occurrences), *descendre* 'to drop' (10 occurrences), *passer* 'to move' (5 occurrences), *dépasser* 'to exceed' (4 occurrences), and *atteindre* 'to reach' (3 occurrences), and the following prepositional phrases: *le seuil de, la barre de,* and, to a lesser extent, *le cap de,* all of which have the meaning of 'the mark' (e.g., *L'inflation dans la zone Euro devrait* ***retomber sous la barre des 6%*** 'Inflation in the Eurozone should fall below the 6% mark' and *M. Issing a lui aussi confirmé que l'inflation devait 'reculer progressivement' pour* ***repasser*** *dans les prochains mois* ***sous le seuil des 2%*** 'Mr. Issing also confirmed that inflation should gradually fall in the next few months to just under the 2% mark').

Table 4: Word frequencies for the ten most frequent verbs occurring with % + preposition de *+ % + preposition* à *+ %*

Rank	*Verb*	*Frequency*
1	*Passer* 'to move'	48
2	*Baisser* 'to drop'	21
3	*Remonter* 'to go up'	19
4	*Porter* 'to move'	17
5	*Revenir* 'to go back'	16
6	*Ramener* 'to reduce'	9
7	*Être* 'to be'	4
8	*Bondir* 'to jump'	2
9	*Reculer* 'to decrease'	2
10	*Aller* 'to go'	2

Verb + preposition à *category*

This category reveals some obvious similarities with the Verb + preposition *de* category. Here also, a limited number of verbs are used to express a percentage with the preposition *à*, the most frequently occurring being *fixer, s'établir, remonter*, and *tomber*. Examples include: *Le taux d'inflation **s'est établi à 2,6%** en décembre 2000* 'The inflation rate was set at 2.6% in December 2000' and *La croissance dans la zone Euro va **tomber à 2%** en 99* 'Economic growth in the Eurozone will fall to 2% in 1999'. Although the total number of occurrences of each of these combinations is small, they nevertheless provide a useful resource for the learner.

Table 5: Word frequencies for the ten most frequent verbs occurring with the percentage symbol (preposition à)

Rank	Verb	Frequency in corpus	Frequency with %
1	*Fixer* 'to set'	329	38
2	*S'établir* 'to be set'	70	37
3	*Remonter* 'to go up'	280	35
4	*Tomber* 'to fall'	332	31
5	*Ramener* 'to reduce'	92	31
6	*Porter* 'to move'	330	23
7	*Se situer* 'to be'	110	15
8	*Passer* 'to move'	652	13
9	*Maintenir* 'to remain'	265	11
10	*S'élever* 'to increase'	42	11

Another similarity with the preposition *de* category is the discovery that *à* is also used differently depending on whether it is placed before or after the percentage symbol. The collocates are not as limited as in the case of *de*, but nevertheless clearly show that when *à* is used directly before the percentage symbol (1,237 occurrences), this usually indicates the end of a sentence, with the most frequent collocate third to the right of the search word being the full stop, for example, *Le niveau des Fed funds est actuellement **fixé à 5%.*** 'The European Development Fund is fixed at 5%'. When *à* is used directly after the percentage symbol (370 occurrences), it is

most frequently preceded by the preposition *de* in the combination mentioned in the Verb + preposition *de* category of: Verb + *de* + % + *à* + %. Another pattern that emerges from this category is the use of the verb *être*, followed by the preposition *à* and preceded either by the adjectives *inférieur(e)* 'lower' or *supérieur(e)* 'higher' in a comparative clause such as *Le taux de chômage **devrait être légèrement inférieur à 9%** de la population active* 'The unemployment rate should be marginally lower than 9%'.

Applications

These results can be exploited in language-learning materials in a number of ways. Tim Johns is well known as one of the first researchers/practitioners to publish exercises which he had created relating to the use of certain lexical items based on concordance results (Tim Johns' Data-driven Learning Page), and these can serve as a useful model for teachers wishing to produce such exercises (see Johns 2003). In the case of this study, they would be particularly suitable in the context of a course in French for business, for example in support of the task of preparing an account of the state of a company or of the economy of a French-speaking country based on documentation provided to the learners. Table 6 is just one example of the type of exercise that could be created based on Johns's framework for concordances. In some cases in Tables 6 and 7 below the full text is consulted and words included in the concordance printout to ensure that the example will make sense to the reader. Translations of Tables 6 and 7 are provided in the Appendix.

Table 6: The top four most frequent verbs used with the percentage symbol: a sample

les banquiers centraux? L'inflation **a atteint 3,4%** en glissement annuel en mai.
ire. La croissance Européenne **devrait atteindre 3%**, ou plus, en 2000 et 2001.
l'objectif officiel du gouvernement, qui **est d'atteindre 70%** des paiements en Euros avant

Le produit intérieur brut américain **a progressé de 5,6%** en rythme annuel au quatrième
Le produit intérieur brut de la zone **pourrait progresser de 3%** en 2000. Dans
L'indice CAC 40 a clôturé **en progression de 1,33%** à 3 942,66 points. Trois autres pla

Dans les années 80, nos exportations **augmentaient de 10%** par an. Dix ans plus tard, nous
Les salaires après impôts devraient **augmenter de plus de 10%** cette année, explique-t-on
Les prix à la production industrielle **ont augmenté de 0,2%** dans la zone Euro en août par

Les prix à la production dans la zone Euro **ont baissé de 0,46%** en décembre 2000, apr
Depuis le début de l'année, Tokyo s'inscrit **en baisse de 5,07%** alors qu'il availt longt
L'inflation, de 7,3%, en glissement annuel en 1996, **baisse à 2,6%** en 1999. Le différentiel

Exercises such as these could of course be included in course-books. Indeed it is surprising that corpora are not commonly used as the source of examples in coursebooks, not even in English where there is no shortage of suitable corpora.

In a number of projects currently being developed in the use of corpora in language learning, the development of online worksheets on a larger scale is planned. In one project in which the authors of this article are involved, it is envisaged that learners would select the area of language use they wished to explore, for example the expression of percentages in French, then select from a variety of meanings they wished to convey, such as rates rising, falling, changing, etc., and then access concordance results pre-selected for them for the item in question. Thus under the heading 'Changing from one percentage to another', the following results (see Table 7) could be presented, chosen from 120 randomly selected occurrences of the structure '*de* + % + *à* + %'. The aim of projects such as this would ideally involve providing not just isolated examples, but an online equivalent to the coursebook in a format which would make it possible to provide a much more extensive range of examples of actual language use.

Table 7: Changing from one percentage to another

Stupeur: en un mois le taux **avait bondi de 2,9% à 4,2%** ! Après sept années d'inflation ma
Le taux de l'obligation assimilable du Trésor français à dix ans **a grimpé de 3,73% à 3,89%**.
Les taux d'intérêt à 10 ans en France **sont remontés jeudi de 4,73% à 4,92%**. Des risques de

Les taux montent quand le cours des titres baisse -, **passant de 5,79% à 5,91%.** Si, comme le

L'inflation reste, elle, presque étale, puisque elle **est passée de 1% à seulement 1,2%.** Aux

En Italie, la taxation des enterprises **recule de 37% à 19%** dès lors que leurs bénéfices sont

l'obligation assimilable du Trésor français (OAT) à dix ans **est remonté de 4,25% à 4,42%.**

Le taux de l'emprunt d'Etat à 30 ans **est revenu de 6,15% à 6%.** Cette progression des valeurs

le rendement de l'emprunt à 30 ans **a baissé cette semaine de 6,06% à 5,88%.** Dans ces conditions,

A more ambitious project could involve giving learners direct access to the corpus, together with advice on how best to make use of it. In the mid 1980s, Johns (1986: 161) recommended placing a C in the margin to advise learners to consult a corpus on the use of the item in question. Clearly, the average learner could not be expected to cope with more than 6,000 results, although random surfing of the results might well produce good results. In the case of this investigation, a random selection of 120 results, following Sinclair's recommendation (1991: 84), provided results very similar to those described in this article, but with much smaller numbers involved, thus providing the learner with a more manageable set of results. Learners could also be advised on search words to use in addition to the percentage symbol to gain access to examples of language used to express quantitative information (Table 8). Using appropriately selected nouns such as *taux* in this context is a useful strategy for discovering what verbs are commonly used with that noun.

Table 8: Advice to learners on how to find quantitative expressions

a) Search for nouns relating to the world of finance to see which nouns are most frequently used and in which expressions they recur. For example: *taux* 'rate', *progression* 'increase' and *augmentation* 'increase'.

b) Search for the adjectives used with these nouns to show how higher and lower quantities are expressed. For example: *inférieur/supérieur* '*higher/lower,* maigre* 'slight' and *fort* 'sharp'.

c) What adverbs are used to modify the quantitative expressions? For example: *Légèrement* 'slightly', *plus/moins* 'more/less' and *actuellement* 'currently'.

Applications such as these not only provide a much larger variety of examples than a published coursebook can include, but also allow an inductive learning mode, making the role of learner similar to that of researcher (Johns 1986) and enabling the learner to interact with the teacher on the basis of real examples rather than intuition.

Conclusion

This study reveals that a semi-specialized corpus as small as 1,000,000 words can prove a valuable resource for language learners and teachers, providing a rich variety of examples of use to accompany and complement the coursebook or grammar in semi-specialized areas and, more particularly, making the advantages of corpus consultation available to those learning and teaching languages where huge reference corpora are not available. Educated native speakers of French in general, and journalists from *Le Monde* in particular, may have little to learn from the above analysis, but a learner of French or a non-native speaker charged with presenting quantitative information in written French may well find the examples of considerable help. More interestingly, however, the findings raise at least as many questions as they answer. In particular the question arises as to how the findings of corpora can be made easily accessible to learners on a large scale, in the way that dictionaries, coursebooks, and grammars are at the moment. The truism that technology is at its most successful when the technology disappears is particularly relevant here, for, in the case of corpora and their accessibility to language learners, the technology has not yet disappeared. Corpus consultation remains an activity practised largely by corpus linguists and not by non-specialists, unless they happen to be students of corpus linguists. There are, however, a small number of examples of resources in which corpora and concordancing are made available in contexts where the technology has effectively disappeared. In the web-based version of the complete works of Maupassant, *Maupassant par les Textes* 'Maupassant through texts', for example, the page giving access to the works also contains the possibility *Recherche de mots et citations* 'Search for words and quotations' (Selva 2003). A user completely ignorant of the existence of corpus linguistics could easily find all the occurrences of a word as easily as by using a search engine. It is interesting to note

that the author comments that this online facility is unique to the best of his knowledge, and asks users to contact him if they come across similar sites in French. Developments such as this could help to make data-driven learning a reality, and obviate the need for teachers to be responsible for creating the resources necessary to enable learners to learn inductively by investigating real language use.

Appendix

Table 6: The top four most frequent verbs used with the percentage symbol: a sample

*L'inflation **a atteint** 3,4% en glissement annuel en mai.* 'Inflation reached an annual low of 3.4% in May'.

*La croissance Européenne **devrait atteindre** 3%, ou plus en 2000 et 2001.* 'European economic growth should reach 3% or more in 2000 and 2001'.

*Ce chiffre est encore très loin de l'objectif officiel du gouvernement, qui est **d'atteindre** 70% des paiements en Euros avant la fin de l'année.* 'This figure falls well below the official government target, that is, to achieve 70% of payments in Euros by the end of the year'.

*Le PIB américain **a progressé** de 5,6% en rythme annuel au quatrième trimestre.* 'In the US, the GDP increased at an annual rate of 5.6% in the last quarter'.

*Le PIB de la zone **pourrait progresser** de 3% en 2000.* 'The Eurozone GDP could increase by 3% in 2000'.

*L'indice CAC 40 a clôturé **en progression** de 1,33% à 3 942, 66 points.* 'The CAC 40 index closed at 3,942.66 points, an increase of 1.33%'.

*Dans les années 80, nos exportations **augmentaient de 10%** par an.* 'In the 1980s, exports increased at a rate of 10% per year'.

*Les salaires après impôts **devraient augmenter de plus de 10%** cette année explique-t-on au ministère des finances.* 'Net salaries

should increase by more than 10% this year according to the Department of Finance'.

*Les prix à la production industrielle **ont augmenté de 0,2%** dans la zone Euro en aôut par rapport à juillet, selon les estimations publiées mercredi par Eurostat.* 'According to estimates published on Wednesday by Eurostat, industrial production prices in the Eurozone increased by 0.2% in August'.

*Les prix à la production dans la zone Euro **ont baissé de 0,46%** en décembre 2000, après une augmentation révisée à 0,1% en novembre, a annoncé Lundi Eurostat.* 'According to figures released by Eurostat on Monday, production prices in the Eurozone fell by 0.46% in December 2000. This follows a revised increase of 0.1% in November'.

*Depuis le début de l'année, Tokyo s'inscrit **en baisse de 5,07%** alors qu'il avait longtemps affiché la meilleure performance des grandes places boursières mondiales.* 'While Tokyo has consistently been recognised as the best performer in the world's major stock markets, since the start of the financial year it has registered a drop of 5.07%'.

*L'inflation, de 7,3% en glissement annuel en 1996, **baisse à 2,6%** en 1999.* 'Inflation dropped from an overall annual fall of 7.3% in 1996 to just 2.6% in 1999'.

Table 7: Changing from one percentage to another

*Stupeur: en un mois le taux **avait bondi de 2,9% à 4,2%**.* 'In just one month, inflation has jumped from 2.9% to a whopping 4.2%'.

*Le taux de l'obligation assimilable du Trésor français (OAT) à dix ans **a grimpé de 3,73% à 3,89%**.* 'The 10-year OAT rate has climbed from 3.73% to 3.89%'.

*Les taux d'intérêt à dix ans en France **sont remontés jeudi de 4,73% à 4,92%.*** 'In France, 10-year interest rates went up again on Thursday from 4.73% to 4.92%'.

De son côté, le rendement de l'emprunt d'Etat américain à dix ans s'est tendu − les taux montent quand le cours des titres baisse −

passant de 5,79% à 5,91%. 'In the US, the yield on ten-year government loans has crept from 5.79% to 5.91%. It would appear that the rate increases when the market price falls'.

*L'inflation reste, elle, presque étale, puisque dans le même intervalle elle **est passée de 1% à seulement 1,2%***. 'Inflation has also remained stable. In the same period, it has only gone up from 1% to 1.2%'.

*En Italie, la taxation des entreprises **recule de 37% à 19%** dès lors que leurs bénéfices sont réinvestis*. 'In Italy, company taxation drops from 37% to 19% as soon as companies reinvest their profits'.

*Le rendement de l'obligation assimilable du Trésor français (OAT) à dix ans **est remonté de 4,25% à 4,42%***. 'The yield on ten-year OAT bonds went up from 4.25% to 4.42%'.

*Le taux de l'emprunt d'Etat à 30 ans **est revenu de 6,15% à 6%***. 'The 30-year government loan rate has gone back from 6.15% to 6%'.

*Le rendement de l'emprunt à 30 ans **a baissé cette semaine de 6,06% à 5,88%***. 'The yield on 30-year loans has dropped this week from 6.06% to 5.88%'.

References

Aston, Guy. 1997. Small and large corpora in language learning. *Practical Applications in Language Corpora*, ed. by Barbara Lewandowska-Tomaszczyk and James P. Melia, 51-62. Łódź: Łódź University Press.

Aston, Guy. 2001. Learning with corpora: an overview. *Learning with Corpora*, ed. by Guy Aston, 7-46. Houston: Athelstan.

Balboni, Paolo. 1986. LGP versus LSP: Which way to the razor's edge? *UNESCO ALSED-LSP Newsletter* 9(1): 22, May: 2-8.

Blanche-Benveniste, Claire. 2001a. Le recueil d'énoncés d'enfants: enregistrements et transcriptions. *Recherches sur le français parlé* 16, 11-37. Aix-en-Provence: Publications de l'Université de Provence.

Blanche-Benveniste, Claire. 2001b. Grammaticalisation d'un terme de lieu : *Quelque part* et *mis à part*. *Recherches sur le français*

parlé 16, 83-101. Aix-en-Provence: Publications de l'Université de Provence.

Cobb, Thomas. 1997. Is there any measurable learning from hands on concordancing? *System* 25(3): 301-315.

Dodd, Bill. 1997. Exploiting a corpus of written German for advanced language learning. *Teaching and Language Corpora*, ed. by Anne Wichmann, Steven Fligelstone, Tony McEnery, and Gerry Knowles, 131-145. London and New York: Longman.

Firth, John R., ed. 1957. *Papers in Linguistics*. London: Oxford University Press.

Grevisse, Maurice. 1997. *Le Bon Usage: Grammaire Française*, ed. by André Goosse. 13th ed. Paris: Duculot.

Habert, Benoît. 2000. Des Corpus representatifs: de quoi, pour quoi, comment? *Linguistique sur Corpus: Etudes et Refléxions*, ed. by Mireille Bilger, 11-58. Perpignan: Presses Universitaires de Perpignan, Cahiers de l'Université de Perpignan.

Hutchinson, Tom and Alan Waters. 1988. ESP at the crossroads. *Episodes in ESP*, ed. by John M. Swales, 177-185. London: Prentice Hall.

Johns, Tim. 2003. Tim Johns Data-driven Learning Page. http://web.bham.ac.uk/johnstf/timconc.htm. Accessed 17 June 2003.

Johns, Tim. 1986. Micro-Concord: a language learner's research tool. *System* 14 (2): 151-162.

Kennedy, Claire and Tiziana Micelli. 2001. An evaluation of intermediate students' approaches to corpus investigation. *Language Learning and Technology* 5(3): 77-90. Available at http://llt.msu.edu/vol5num3/kennedy/default.html. Accessed April, 2003.

Lawson, Aaron. 2001. Rethinking French grammar for pedagogy: the contribution of spoken corpora. *Corpus Linguistics in North America. Selections from the 1999 Symposium*, ed. by Rita C. Simpson and John M. Swales, 179-194. Ann Arbor: University of Michigan Press.

Partington, Alan. 1998. *Patterns and Meanings. Using Corpora for English Language Research and Teaching.* Amsterdam: John Benjamins.

Sager, Juan C., David Dungworth, and Peter F. McDonald. 1980. *English Special Languages: Principles and Practice in Science and Technology.* Wiesbaden: Athenaion.

Scott, Michael. 1999. *Wordsmith Tools*. Software. Oxford: Oxford University Press.

Selva, Thierry. 2003. *Maupassant par les textes* http://www.maupassant.free.fr (Accessed 17 June 2003).

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Swales, John M. 1990. *Genre Analysis: English in Academic and Research Settings*. New York: Cambridge University Press.

Tomlinson, Brian. (ed.) 1998. *Materials Development in Language Teaching*. Cambridge: Cambridge University Press.

Varantola, Krista. 1986. Special language and general language: Linguistic and didactic aspects. *UNESCO ALSED-LSP Newsletter* 9(2): 23, December: 10-19.