S0958344000000422a 13/11/00 4:42 pm Page 1

ReCALL **12** (2): 157–169. Printed in the United Kingdom

© 2000 Cambridge University Press

Concordancing and comprehension: preliminary observations on using concordance output to predict pitfalls

MARIE-MADELEINE KENNING
School of Language, Linguistics and Translation Studies, University of East Anglia, Norwich NR4 7TJ, UK

Abstract

This paper presents an innovative application of concordancing to the study of comprehension based on the hypothesis that the presence of noise in raw concordance data brings to light the existence of strings liable to cause unsuspected comprehension problems to language learners. The paper explains how the hypothesis was arrived at before proceeding to describe and discuss an experiment specially designed so as to test its validity. In the course of this experiment eight post A-level students and two native speakers/teachers of French were observed processing the raw output of a concordance of the string *bien que*. Details of the concordance data are given, broken down in broad categories (concessive conjunction *bien que*, resultative conjunction *si bien que*, adverb + completive, etc.). The findings are reported in three stages corresponding to those of the experiment: a pilot study involving two of the students, a repeat of the experiment with the two native speakers, and a second learner study with a larger, more representative sample of six students. Between them the results are found to provide empirical evidence in support of the hypothesis. The main outcome is that contrary to the native speakers, none of the post A-level students was able to interpret all the citations correctly. Other significant results are that difficulties were spread unevenly across citations and were more common among those with lower grades. The implications for language pedagogy are then examined. The paper concludes that concordances can be helpful in predicting potential pitfalls and ends with some suggestions for future research.

1 Introduction

This paper presents the findings of a small-scale experiment involving the use of concordance data in the study of language processing. The aim of the experiment was to explore an hypothesis prompted by previous work on concordances: that the presence of noise in raw concordance data brings to light the existence of strings liable to cause unsuspected comprehension problems to language learners.

In common with the data-driven learning (DDL) approach advocated by Tim Johns (Johns, 1991a, 1991b), the experiment only employed technology as a tool for gathering data, using a keyword-in-context (KWIC) concordancer and a corpus of texts to generate a list of citations. The part played by technology was nevertheless essential, since to collect such data without a concordance package would be prohibitively time-consuming.

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 2

Unlike DDL, and in contrast with most of the literature on concordances within an educational context (see items in Johns, 1992b; also Guillot, forthcoming), the experiment did not focus on the exploitation of citations for teaching purposes, whether on-line or off-line; as mentioned above, it was concerned with the contribution that concordance output might make to the identification of comprehension pitfalls; as such, it is believed to represent an innovative application of concordancing. The ultimate intention, however, was to enhance the quality and effectiveness of language learning and, as shown below, the fresh insights into comprehension problems gained through this kind of research have clear implications for language teaching.

The focus of attention of the experiment revolved around the parsing of raw concordance data produced by searching for the string *bien que*. The subjects were post A-level students of French and the main part of the investigation was carried out in two stages. First a pilot study was undertaken to test the methodology and ascertain the suitability of the data and the fruitfulness of the approach. The experiment was then repeated with a larger and more representative group of students.

The account that follows begins with a description of the antecedents of the study. This is followed by a section on methodology, which in turn leads to the presentation and discussion of the results of each of the two stages. The paper ends with an assessment of the approach and the pedagogical implications of the findings.

2 Motivation for the study

The inspiration for the study was an incident that took place in the course of a project involving the compilation of concordances of *ce que* where *que* functioned as a relative pronoun. As described in Kenning (forthcoming), that particular project had two related aims: to establish the relative frequency of relative pronouns with a view to informing a first year post A-level grammar course, and to supply instances of actual use for inclusion in illustrative or practice materials. The concordances were produced using *Multiconcord*, a multilingual parallel concordancer developed with LINGUA funding by a team headed by Francine Roussel of the University of Nancy[1]. They were based on a subset of the corpus of parallel texts assembled by the team, consisting of pairs of French and English files representing a variety of genres: three chapters of a scientific book, a work of popularisation of academic writing and two works of fiction. These are the kind of texts that one would expect a post A-level student to have to deal with, so that, while no claim can be made about the representativity of the corpus in relation to French, it does provide a valid basis for pedagogically oriented investigations. In two cases (the academic text and one of the fictional works) the French text was the source text, while in the other two the French was a translation from English. The project employed a research assistant with a degree in French whose task included cleaning out the data, counting the number of relevant citations and gathering interesting examples.

To my dismay, the citations which my research assistant presented to me proved to contain not only instances of the relative pronoun *que*, but also a disturbing number of various other uses of *que*. The search for *ce que* had resulted in output showing sentences such as

[1] For a review of *Multiconcord* see St. John & Chattle (1998).

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 3

Finalement, en 1915, il proposa ce que nous appelons maintenant la théorie de la Relativité Générale
(Finally, in 1915, he proposed what we now call the general theory of relativity)

and

Les deux rives du Bosphore se ressemblent comme les versants d'une vallée, ce qu'il était jadis
(The two banks of the Bosphorus resemble each other like the slopes of a valley, which is what it once was)

but also, *inter alia*,

Je ne comprends pas ce que vous voulez dire
(I don't know what you mean)

Ce qu'on va me trouver courageuse!
(How brave they'll think me at home!)

Est-ce qu'il collectionne les papillons?
(Does he collect butterflies?)

Qu'est-ce que ça peut bien être?
(What is it?)

and

Une autre [...] y vécut plusieurs mois comme une 'vraie sauvage' jusqu'à ce qu'un navire génois la recueille par hasard.
(Another woman [...] lived there for several months until a Genoese ship rescued her by chance).

Some of the rogue occurrences had been taken out, but others were still there. It was clear that I had underestimated the difficulty of cleaning out data. I realised at that point that I had little idea of the strategies involved in the cleaning out process. Was it a question of selecting items or rejecting intruders? What sort of cues did I myself use? How did these vary from citation to citation?

Looking at the last three citations one's intuition is that what enables native and proficient speakers to sort out true instances of *ce que* from cases in which this string forms part of a larger unit is the presence of elements such as *qu'est-*. It seems likely that such speakers have separate entries for *ce que, est-ce que, qu'est-ce que, jusqu'à ce que* in their mental lexicon facilitating their recognition. What is interesting is that these larger units are so much apprehended as single groups that their appearance in the concordance comes as a surprise. Briscoe (1987:22–23) notes a similar effect in the face of ambiguity:

It is striking that most ambiguity, whether local or global, is simply not noticed by

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 4

the majority of language users and this testifies to the efficiency and robustness of their mechanisms for disambiguation.

Distinguishing between relative clauses, indirect questions and exclamations in which *ce que* functions as degree adverb is a complex task requiring competence and skill, as shown by my assistant's inability to spot all the intruders. This leads to the hypothesis that raw untagged data may be of interest from a language acquisition perspective and used to gain new insights into the parsing difficulties facing naive, less proficient learners.

The empirical study described below is a practical application specifically designed to test this hypothesis. It is based on the premise that output containing 'noise' can incidentally reveal potential areas of confusion outside a native speaker's awareness (and thereby eventually contribute to better teaching). The materials used in the study involve a concordance of *bien que*, such as might be undertaken in order to find authentic examples of the concessive conjunction *bien que* (although)[2]. Again, the unsorted results are a mixed bag comprising the kind of sentences that form the object of the search, as well as a number of sentences in which the sequence *bien que* does not constitute a unit but is instead the surface manifestation of other selections.

In the light of many years of teaching post A-level French, this data seemed suitable for my purpose. In view of the high profile of the conjunction *bien que* through, notably, its inclusion in exercises on the subjunctive, I expected tokens of the conjunction to be recognised by students. I was less certain that they would correctly parse and understand other citations. I knew they had come across at least some of the choices underlying the generation of the *bien que* string, but thought the stress laid on the conjunction likely to have made it so prominent in their minds as to lead them to overlook other possibilities.

In effect raw output containing pairs of words that have multiple syntactic/semantic functions can be seen as putting students in a situation reminiscent of garden path studies. Garden paths are sentences in which the problem of disambiguation, which normally passes unnoticed, "does surface phenomenologically", most cases being "local ambiguities which are typically misconstrued by readers when they are presented out of context, without punctuation" (Briscoe, 1987:23). A made up example relevant to our investigation would be

Il comprend bien qu'on ne lui en parle pas

where *bien que* may be interpreted either as the concessive conjunction (he understands although no-one mentions it to him) or as the juxtaposition of the adverb *bien* and the conjunction *que* (he fully understands that no-one mentions it to him). Another, perhaps more convincing example, would be

Elle parle si bien qu'on est forcé de l'écouter

which could mean either that she speaks so well that you are spellbound listening to her,

[2] For an account of the concessive uses of *bien que* see Morel (1996).

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 5

or that she will insist on speaking and you have no choice but to listen. While in either case the ambiguity cannot be resolved morphosyntactically in the sentence as it stands, in normal circumstances ambiguity would be reduced by the expectations created by the context, and dispelled by the use of a comma or the intonation. Undeniably, as Briscoe (1987:49) indeed acknowledges,

> garden path examples are artificial, in the sense that they are created by distorting or removing some of the information normally available to listeners and readers.

They are nevertheless useful for gaining insights into the kind of knowledge deployed by readers/listeners and the strategies which lead to particular interpretations, and for this reason have attracted a great deal of attention.

The point being made here is that while the *bien que* concordance differs from garden path material in presenting authentic unambiguous data, the parsing tasks involved present two important similarities: the first is that in both cases subjects find themselves dealing with multiple possible analyses – of a single sentence in the case of garden paths, of the same string across sentences in the concordance; the second point of resemblance is that in common with people puzzling over garden paths, post A-level students find themselves operating with fewer clues than 'normal' (taking a native speaker's situation as the norm). Not because the clues are not there, as in genuine garden paths, but because of their reduced ability to pick them up. The obstacle is one of accessibility rather than availability, but the effect is close enough to warrant the analogy.

Interestingly, native speakers are less aware of what might lead to the occurrence of *bien que* than of the ambiguity of polysemous items. Even when alerted to the fact that the co-occurrence of *bien* and *que* may be something else than a token of the subordinating conjunction *bien que*, they find it difficult to list the full range of possible sources. Taking *que* as an example, I explained to two colleagues who are native speakers of French how *Multiconcord* would pick up all instances of *que* irrespective of their function, so that, if one was interested in, say, the use of *que* as relative pronoun, one would have to remove manually all other citations. I then asked them whether they thought one would encounter similar problems when looking for the conjunction *bien que* and noted and timed their responses. Both answered yes but, as shown by Table 1, produced incomplete lists and took some time to do so.

Why some possibilities spring to mind while others do not, and whether certain patterns would emerge from a larger sample, merits further investigation. What is clear is

Table 1. *Native speakers' awareness of origins of* bien que

Subject 1	Subject 2
adverb + *que* (5sec)	noun + *que* (immediate)
aussi bien que (20sec)	adverb + *que* (40 sec)
si bien que - comparative (60sec)	
noun + *que* (85sec)	
si bien que - conj (120sec)	

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 6

that native speakers find it difficult to think of appropriate contexts for the occurrence of *bien que*. This is an interesting finding because it does not seem unreasonable to assume that problems associated with producing appropriate contexts turn into parsing difficulties in the case of less competent language users. In other words, it is tempting to interpret the native speakers' imperfect responses as evidence of an intrinsic complexity which manifests itself in others in ways that vary according to the level of proficiency of the subject. Equally, the time taken to bring to consciousness what is only a subset of possible contexts may be taken as indicating lack of awareness of the potentially confusing nature of *bien que*. This would obviously lend support to our contention that concordancers are capable of revealing unsuspected problem areas, should *bien que* prove troublesome.

3 Methodology

The search for occurrences of the sequence *bien que* in the files mentioned above resulted in a total of 72 hits, all of which were included in the materials presented to the students. These materials consisted of four sets of numbered French sentences obtained by removing the English version and text references from the output. Contrary to what is common practice with concordance-based handouts, the appearance was that of an ordinary list of sentences with nothing marking out the search phrase. To highlight *bien que* would have further distorted[3] the interpretation process relatively to ordinary reading and would have been misleading in citations other than the conjunction.

Evidence was gathered by means of a think aloud protocol in a tutorial setting. The materials were handed out to students as they sat down, with a request to work out aloud the meaning of the segment containing *bien que*. What they did, how long they took over it, how much of the sentence they took in, whether they translated or explained, was left up to them. In practice all the students elected to translate, although some also made an occasional metalinguistic comment.

Three reasons lay behind the choice of a think aloud protocol: it was less time consuming for the students; it allowed me to eavesdrop on the interpretation process and jot down any interesting feature (in addition to the responses themselves); and it enabled me to provide feedback as soon as the task was completed as well as to explore possible explanations for mistakes with the first round still fresh in the student's mind.

Since the most significant variations in the meaning of the string relate to differences in the function of the words or their immediate constituent(s), this was used as the basis for classifying the concordance data. This yielded the following seven categories:

- concessive conjunction *bien que*
- resultative conjunction *si bien que*
- adverb + completive
- adverb within *ne ... que*
- conjunction *ou bien* + completive
- conjunction *aussi bien que*

[3] 'further' because it is clear that processing a concordance under experimental conditions does not constitute ordinary reading. One should try, however, to minimise differences.

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 7

Table 2. *Occurrences of* bien que *in concordance data*

file	conj. *bien que*	conj. *si bien que*	adv. + completive	adv. within *ne … que*	conj. *ou bien que* + completive	conj. *aussi bien que*	adv. within comp. struct.	total
1(F/E)	9					3	1	13
2(F/E)			11	1			1	13
3(E/F)	8	1			2(*2)	1		14
4(E/F)	2	4	23				3	32
Total	19	5	34	1	4	4	5	72

- adverb within comparative structure

The above list should not be taken to constitute an exhaustive inventory of all possible occurrences of the segment *bien que*. Our purpose was not to provide a comprehensive taxonomy but simply to devise a convenient framework for analysing the response data. This pragmatic emphasis means that the list reflects the nature and limited size of the corpus. Hence the omission, *inter alia*, of the noun + *que* category (*Il ne faut pas oublier tout le bien qu'il a fait*, one must not forget all the good he did) mentioned by both our speakers.

The number of occurrences in each category is displayed in Table 2.

The primary aim of the response data analysis was to establish whether subjects had an adequate sense of the function (and basic meaning) of the *bien que* string. This was assessed after the interview by classifying responses in terms of the categories of Table 2, since these encapsulate the kind of major distinctions we were interested in. In so doing I was led to mark a number of flawed responses as satisfactory, but took note of these minor blemishes and comment upon them below, as well as upon various other noteworthy features.

4 The pilot study

The pilot study involved two post A-level students referred to below as Subjects 1 and 2. Both subjects had achieved an A in the A-level examination. Subject 1 was at the end of the first year of a French and Italian degree, Subject 2 was a 2nd year History student with some experience of reading French historical texts. In terms of level and exposure to French, the subjects belonged therefore to the upper part of the ability range, yet were still sufficiently representative to carry out a preliminary assessment of our hypothesis.

4.1 Concessive conjunction bien que

Both subjects correctly spotted all the tokens of the conjunction *bien que* occurring in the data. However, some hesitancy was noted in four cases. Subject 1 had difficulty with

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 8

Bien que ces notions qui nous semblent justes fonctionnent lorsqu' il s'agit de pommes ou de planètes qui se meuvent relativement lentement, elles ne sont plus du tout valables dans le cas d' objets se déplaçant à la vitesse de la lumière ou presque.
(Although our apparently commonsense notions work well when dealing with things like apples, or planets that travel comparatively slowly, they don't work at all for things moving at or near the speed of light.)

with

La masse du Soleil incurve l'espace-temps de telle sorte que, bien que la Terre suive une trajectoire droite dans un espace-temps à quatre dimensions, elle nous apparaît comme se mouvant le long d'une orbite dans un espace à trois dimensions.
(The mass of the sun curves space-time in such a way that although the earth follows a straight path in four-dimensional space-time, it appears to us to move along a circular orbit in three-dimensional space.)

and with

Bien que longtemps discutée, une relation de cause à effet ne paraît pas étroite ni absolue.
(Although it has been discussed for a long time, there does not appear to have been a definite or absolute relationship of cause and effect.)

while Subject 2 had problems with

Bien que sans aucune commune mesure avec les fleuves chinois ou l'Amazone, ceux de l'Europe prolongent jusqu'au coeur du continent les rythmes de la navigation maritime: le Danube jusqu'à Budapest et Vienne, le Rhin jusqu'à Strasbourg.
(Although they cannot be compared to the Chinese rivers or the Amazon, those of Europe prolong the rhythms of maritime navigation into the heart of the continent: on the Danube, as far as Budapest and Vienna, and on the Rhine as far as Strasbourg.)

In all four cases the sentence was felt to be complex, either because of embedded clauses or as a result of the omission of the verb or auxiliary in the concessive clause. In the case of the last sentence but one the problem arose from an initial incorrect linking of bien and longtemps.

4.2 Resultative conjunction si bien que

These posed no problem to Subject 2. They were recognised as forming a unit by Subject 1 but the latter experienced some difficulty in finding a good English equivalent. Typically, faced with

Selon la seconde loi de Newton, ces deux effets s'annuleront exactement l'un l'autre, si bien que l'accélération sera la même dans tous les cas.

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 9

(According to Newton's second law, these two effects will exactly cancel each other, so the acceleration will be the same in all cases.)

she first proposed 'provided that', then 'so long as' before being guided to 'so that' by the presence of the future tense in the subordinate clause.

4.3 Adverb + completive

This pattern, which as shown by the table above was the most common overall, caused some problems to both subjects. In all 34 cases, *bien* acted as intensifier, but its precise meaning varied with that of the qualified item. These were, in order of decreasing frequency: *vouloir* (9), *savoir* (6), *croire* (5), *peut-être* (3), *sembler* (3), *sentir* (2), *aimer* (1), *deviner* (1), *espérer* (1), *falloir* (1), *supposer* (1), *voir* (1).

Subject 1 had fewer difficulties than Subject 2, failing only once to pick up the intensive function of *bien*. The one exception was a sentence containing the string *tout en sachant fort bien que*, where she mistook *bien + que* for the concessive conjunction. The sentence was also misconstrued by Subject 2 who translated *fort* as strongly and then got stuck. The problem here seems to have been the subjects' lack of familiarity with the use of *fort* as adverb and it may have been aggravated, as one of them suggested, by the fact that *fort* appeared at the end of a line, with *bien* (and *que*) on the next line.

Subject 2 had difficulty with a further three sentences, settling for 'although' in the case of *devina bien que* where she did not know the meaning of *deviner*, going for 'although' but then correcting herself in the case of *savais bien que* (at the beginning of a fairly complex sentence), and proposing 'although', then 'while', before eventually giving up when faced with *sentait bien que*. Again she saw her difficulties as linked to the overall complexity of the sentences.

4.4 Adverb within **ne ... que**

In this case *bien* functioned as an adverb of manner qualifying *voir*. Both subjects interpreted the sentence correctly.

4.5 Conjunction **ou bien** *+ completive*

Both subjects had problems with the first case of *ou bien que ... ou bien que*, taking a long time over it. In both cases the subject started by parsing the string as containing a concessive conjunction before realising that this did not work and deduced the meaning 'either .. or' from the rest of the sentence. Neither then had any problem with the next case, which followed immediately.

4.6 Conjunction **aussi bien que**

All four citations of *aussi bien que* were parsed correctly, but Subject 1 was hampered by the overall complexity of the first citation and Subject 2 mistranslated one case as 'more than'.

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 10

4.7 Adverb within comparative structure

These comprised three straightforward comparisons (e.g. *nageais moins bien que*) and two occurrences of *tant bien que mal*. Neither subject misconstrued the comparisons, although Subject 2's first response (involving a comparison with *aussi*) was rather slow. *Tant bien que mal* posed no problem to Subject 1 who noticed the opposition *bien/mal*. Subject 2 did not know the expression: her first encounter with it led to an initial misanalysis in terms of 'although', which she then rectified, spotting the link between *tant* and *que*. The second occurrence was correctly parsed without any progress being made on what the expression meant.

5 Native speakers and the parsing of *bien* + *que*

For comparison purposes, the experiment was repeated with two native speakers, using the same materials. Unlike the subjects in the pilot study, the native speakers had no difficulty in interpreting the citations. The speed with which they did this was strikingly faster than the learners; however, as with the learners, reaction times were affected by overall sentence complexity, with more time being spent on some of the sentences which the learners had found difficult, notably those containing instances of the concessive conjunction placed mid-sentence and/or introducing a verbless clause. More interestingly perhaps, interpretation appeared to take place in a more linear manner, with little evidence of interaction between syntactic and semantic processing. By contrast, as has already been seen, the learners tended to use semantic processing to check their interpretation. Indeed in the case of Subject 2, this represented a deliberate, often successful strategy, which enabled her to achieve equal success with Subject 1.

6 Preliminary conclusions

In so far as the subjects did not interpret all the citations of *bien que* correctly, the pilot study was seen as providing support for regarding this string as a potential source of comprehension problems, and by implication as providing support for the view that untagged output can supply useful material for studying parsing procedures. Since the subjects only produced four cases of incorrect or unsuccessful parsing out of a total of 142 citations, with a further four cases of initially faulty but subsequently remedied interpretations, the evidence was not as strong as anticipated. However this had to be weighed against the above average general ability of the two students, as measured by their A-level results. In addition, the study confirmed both the learners' familiarity with the use of *bien que* as a concessive conjunction, and the prediction that they would be biased towards this use. Here the evidence was clear, with the conjunction being used as default in seven out of the eight problem cases.

In the light of these findings, it was decided to replicate the study with a larger sample of six first-year students enrolled on a French post A-level course. The A-level grades of these students included one A (subject A), two Bs (subjects B and C), two Cs (subjects D and E) and one D (subject F), a pattern which is representative of the distribution found in the institution.

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 11

7 The second study

The second study took place towards the end of the first semester. Table 3 shows the number of satisfactory responses given by each student, together with their most common incorrect responses. Please note that the responses mentioned were selected on the basis of frequency and that in most cases students also provided other responses, or admitted that they did not know.

Again, there were noticeable differences between students in terms of reaction times, processing strategies and response quality, which Table 3 does not capture. For example, besides identifying all the occurrences of *bien* as intensifier, subject A provided superb, varied translations demonstrating an obvious flair for language. Also worth mentioning is the plausibility of B's (wrong) rendering of *bien que* and *si bien que* in the contexts involved, to the point of representing a more likely option in one case than the correct translation!

The picture that emerges from the analysis of this new set of data corroborates the findings of the pilot study. First and foremost, it confirms the existence of problems in interpreting *bien + que*. These problems have various origins. They may be due to lack of awareness of a particular meaning, function or unit, as with *si bien que*, or *tant bien que mal*. Alternatively they may be traceable to something in the verbal context. This may be a localised difficulty caused by a specific item, like not knowing *deviner*, or mistaking it for *devenir*, or a general feature of the text such as a high degree of syntactic complexity.

Table 3. *Number of satisfactory responses*

	conj. *bien que*	conj. *si bien que*	adv. + completive	adv. within *ne ... que*	conj. *ou bien* + completive	conj. *aussi bien que*	adv. within comp. struct.	total
	19	5	34	1	4	4	5	72
A	19	0 ‘even if’	34	1	4	4	3	65
B	16 ‘because’	4 ‘because’	29	1	0	3 ‘although’	3	56
C	19	0 ‘if although’	3 ‘although’	1	0 ‘or although’	0 ‘also though’	0 ‘less/also though’	23
C	19	0 ‘if although’	3 ‘although’	1	0 ‘or although’	0 ‘also though’	0 ‘less/also though’	23
D	18	0 ‘although’	26	1	0 ‘although’	3 ‘although’	3	51
E	13	0 ‘even if’	31	0	0	4	3	51
F	0 ‘just as/ that’	0 ‘only if’	28	0	0	2	3	33

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 12

The data also confirms a general, though not universal, tendency to use 'although' as default. Some learners are more prone to do this than others. Subject F did not, but was not likely to, given his inability to identify tokens of the conjunction. Neither did Subject E, in contrast to Subject C, who used 'although' erroneously either on its own, or in combination with some other item like 'if', or 'also'. It should be acknowledged here that the temptation to use 'although' may have been exacerbated by the nature of the task which, in inviting the subjects to focus on *bien* + *que*, may have unwittingly pointed them in the direction of the conjunction. Nevertheless the trend was prevalent enough to support the contention that the stress put on the concessive conjunction is liable to interfere with comprehension.

Finally, in providing further evidence that raw concordance data can be used to disclose and document comprehension difficulties, the study provides validation for the approach and opens up a new type of application for concordance work.

8 Conclusion

The findings reported above have a number of implications for the teaching of French. It is obviously essential to try and plug the more glaring gaps in the students' knowledge revealed by the study, notably the striking absence from their linguistic repertoire of *si bien que, ou bien,* and *tant bien que mal*, which are all relatively common strings immediately identifiable for native speakers. A high priority should also be given to the intensifying function of *bien* so that it is recognised in all contexts, including after *peut-être*. Finally, there are grounds for considering putting less emphasis on the concessive conjunction in A-level teaching, or at least for making learners aware that *bien* and *que* can co-occur for other reasons.

It goes without saying that these various forms of remedial action might usefully employ concordance-based materials. Indeed, can there be, for instance, a more effective means of developing student awareness of the broad range of contexts in which *bien* acts as an intensifier, and of the varied ways in which this function is conveyed in English, than through sets of parallel sentences drawn from concordance data? The extent to which this kind or remedial action can be accommodated in the curriculum will of course depend on the objectives and priorities of the course. In any case, given the premium generally placed on class time, it would seem best to adopt the pedagogical approach described and illustrated in Johns (1991b) and, except in the case of very advanced or able students, use prepared handouts that provide controlled sequences of activities and exercises, rather than concordance printouts. Because working with this type of materials requires, as well as stimulates, inductive strategies, students would need some in-class initiation, but the handouts would be intended and designed for out-of-class use.

More generally, in focusing on comprehension the study invites us to reflect on the place given to its development in language teaching. Have we currently got the balance right between comprehension and production? Should we perhaps re-examine our priorities in the light of technological advances that are revolutionising access to target language texts and changing the way we communicate? If life is to put a growing premium on information gathering, and if Hawkins is right in suggesting that "increasingly, 'bilingual' dialogue and correspondence, each side using its own language, may become

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press

S0958344000000422a 13/11/00 4:42 pm Page 13

the rule" (Hawkins 1994:121–122), as already often happens in tandem language learning, then more attention needs to be paid to the fostering of comprehension, and to the detection of the kind of obstacles discussed in this paper.

Bearing in mind that problematic strings are not perceived as such by competent users and that their identification has so far been largely a matter of serendipity, how is the task of collecting such strings to be approached? As teachers we may each be aware of one or two strings that can be parsed in more than one way and might try to pool our intuitions or to use the ubiquity of words like *que* as a starting point for brainstorming sessions. But this is not likely to get us very far. Another possibility would be to build looking for garden path effects into the objectives of a larger project. Again this is not likely to be very productive: resources will have to be spread, other objectives will probably take precedence, and the concordances may fail to produce much relevant data. The best way forward would seem to be to set up a collaborative project specifically aimed at gathering evidence on garden path effects in one or several languages that capitalises on the results of past and current research. The project would adopt a multi-pronged approach comprising analyses of published papers, requests for access to concordance data, the distribution of questionnaires to researchers and the establishment of a network for the transmission of new potentially relevant findings.

To verify the impact of the strings on comprehension one might employ the methodology outlined above or, if the overall aim is to collate information on as many potential pitfalls as possible, devise a simpler, less time-consuming procedure. If, on the other hand, the project is more concerned with studying comprehension processes, one would require a more sophisticated methodology. Either way there is a lot of interesting work to be done.

References

Briscoe, E. J. (1987) *Modelling human speech comprehension: a computational approach*. Chichester: Ellis Horwood

Guillot, M.-N. (forthcoming) Le travail sur corpus électronique au service de l'apprentissage de la langue: de la théorie à la pratique. In: Lewis, T. and Rouxeville, A. (eds.), *Technology and the Advanced Language Learner*, London: CILT/AFLS.

Hawkins, E. (1994) Percept before precept. In: King, L. and Boakes. P. (eds.), *Grammar! A conference report*. London: CILT, pp. 109–123.

Johns, T. *Classroom Concordancing/Data-driven learning Bibliography.* http://web.bham.ac.uk/johnst/biblio.htm

Johns, T. (1991a) Should you be persuaded: two examples of data-driven learning. In: Johns, T. and King, P. (eds.), *Classroom Concordancing*. Birmingham: Centre for English Language Studies, University of Birmingham, pp. 1–16.

Johns, T. (1992b) From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In Johns, T. and King, P. (eds.), op. cit., pp. 27–45.

Kenning, M.-M. (forthcoming) Corpus-informed development: parallel concordances and pedagogical grammars. In: Lewis, T. and Rouxeville, A. (eds.), *Technology and the Advanced Language Learner.* London: CILT/AFLS.

Morel, M.-A. (1996) *La concession en français*. Paris: Ophrys.

St.John, E. and Chattle, M. (1998) Multiconcord: the Lingua Multilingual Parallel Concordancer for Windows. *ReCALL Newsletter*, **13**: 7–9.

https://doi.org/10.1017/S0958344000000422 Published online by Cambridge University Press