Reihe:
FREMDSPRACHEN IN LEHRE UND FORSCHUNG (FLF)

Im Auftrag der Ständigen Kommission des Arbeitskreises der Sprachenzentren, Sprachlehrinstitute und Fremdspracheninstitute (AKS)

herausgegeben von

Klaus Vogel und Bernd Voss

Band 31

Coleman, James A., Grotjahn, Rüdiger & Ulrich Raatz (eds.)

University Language Testing and the C-Test

AKS-Verlag, Bochum

2002

Explanation of the Verbal Protocol Mapping Graph

1. Reading aloud is represented by an uninterrupted line. The beginning of each line is marked by ⊤, the end by ⊥. One uninterrupted line with a beginning and an end mark is called a reading cycle.
2. The numbers to the left are referring to the atomic proposition number of each element.
3. Repetition of *one* single item is not marked specifically. Similarly, repetition of an item while the subject is writing it down is not taken into account.
4. Pauses are not marked.
5. Movement to a new reading cycle is performed when sequential reading is discontinued, such as: subject regresses to previous text elements and re-reads, translates etc.; or subject continues reading the text but utters word which does not sequentially follow the last word which was uttered. If *one* item was skipped during reading aloud and the subject uttered the next word after the one which had been skipped, this was not taken as a criterion to start a new reading cycle. If, however, more than one element was skipped, this was taken as an indication to start a new reading cycle.
6. Subject-item interaction during reading aloud is marked as follows: white circle = item was not completed (only item beginning was uttered); black circle = item was completed correctly; crossed circle = item was completed incorrectly.
7. Try 2, Try 3 etc. denote that the subject works through the text once again after having worked on one of the other texts.

* = item discussed in case study
|| = translation
/ = word was skipped

Eight: Psycholinguistics of C-Test taking†

Christine Klein-Braley

One of the most important questions about a language test for the psycholinguist is how far the test elicits authentic language behaviour from the test subject. Indeed, one of the main criticisms of multiple-choice language tests is that putting crosses in boxes has no relationship to real-life language use. One of the questions most frequently asked about C-Tests is whether the behaviour elicited by the mutilated texts can be genuinely viewed as representing a sample of the test takers' general language proficiency. One possible answer to this question is pragmatic: C-Tests have high intercorrelations with many other criteria which are viewed as representing real-life language use: teacher grades, other tests and examinations, pupils' self assessments, oral interviews, essay tests, reading and listening comprehension tests and so on. Therefore, if the other criteria represent real language use, so does the C-Test. For the psycholinguist, however, this statement, while true, is not satisfying. What goes on when a person takes a C-Test?

One way of looking at the behaviour shown by C-Test takers is the investigation of test-taking processes. Feldmann *et al.* (1986; see also Grotjahn, 1986) suggest three possible approaches: *statistical item analysis*, *text linguistic item analysis* and *analysis of individual performance*. The research in Duisburg conducted by Raatz and myself has primarily used the classical statistical techniques of test and item analysis. The Bochum group under the leadership of Grotjahn have concentrated their efforts on investigations of individual performance while subjects are completing C-Tests. This paper will present evidence derived from the linguistic analysis of test-taker responses. A more detailed account of the research summarised here can be found in Klein-Braley 1996 (in Grotjahn, 1996).

1. Investigation of test-taking strategies: think-aloud protocols

The researchers in Bochum (Grotjahn, 1987; Stemmer, 1991, 1992; Feldmann *et al.*, 1986) have chosen to use think-aloud procedures to investigate C-Test taking behaviour. There is no doubt that analysis of these protocols has substantially increased our understanding of what goes on inside the subject during test-taking. The methodology involves asking subjects to verbalise their thoughts while taking the tests. After the test has been completed, the tapes are then replayed to the subject who can answer any questions put by the investigator, clarify remaining uncertainties, or add comments if necessary. But there are problems in using this technique. One major drawback for anyone trained in a quantitative tradition is the *small number of subjects involved.* Feldmann *et al.* (1986) report results for a total

of 20 subjects in all (10 each in Spanish and French). Stemmer's 1991 study reports data for 30 subjects.

Another problem is that the results may be affected by a very considerable amount of *reactivity*. The think-aloud protocols concentrate on cognitive processes, but an interaction between the process of test-taking and the process of commenting on test-taking cannot be excluded.

What seems to me to be most problematic, however, is the question of *task authenticity*. In terms of modelling C-Test performance, the think-aloud experiment is very far from real life. Stemmer's subjects were tested individually and she made great efforts to give them the feeling that they were 'competent collaborators' (Stemmer, 1991: 55). How far the results derived from such sessions can answer the question of what happens when C-Tests are processed under 'normal' conditions - in a large room with other test takers, under pressure of time and the stress of being tested - is still very much open to question. Think-aloud techniques, while informative, are thus not conclusive evidence of what goes on in C-Test processing.

2. Investigation of test-taking strategies: non-reactive approaches

Is there any way of avoiding the problem of reactivity? It seems to me that the other two approaches suggested by Grotjahn offer precisely this opportunity. They involve first an *ex post facto* analysis of the statistical indices of individual items and their interrelationships, with the aim of interpreting these in terms of linguistic phenomena. No evidence from this area is presented here, but see Klein-Braley (1996). Secondly, it is possible to perform an inspection of the test-taking behaviour of the test subjects as evidenced by the test scripts. Again, the aim is to detect the psycholinguistic processes involved in test taking. This is the purpose of the present paper.

3. Studying psycholinguistic processes in test-taking

3.1 The basic data used in this investigation

I base my conclusions on data drawn from over 300 subjects who completed in all a total of 30 different English C-Test texts. Tables 1 and 2 summarise the basic data. Around one third of the group are English 11- to 13-year-olds from Wales Comprehensive School in South Yorkshire.[1] For these subjects no validation data

[1] My thanks to the teachers and pupils who cooperated, in particular to Ms Judith Cole.

Table 1: Basic data German group

Test	Form	n	Mean	SD	P	r_{tt}	r_{DELL}	r_{dict}	r_{tot}
CARS		120	17.91	4.86	71	.85	.69	.73	.72
LITERATURE		180	12.12	3.93	52	.78	.60	.59	.61
CARS	1	23	16.61	4.92	63	.85	.77	.81	.79
	2	21	17.23	5.29	67	.87	.53	.64	.57
	3	22	16.90	4.87	64	.84	.79	.82	.82
	4	25	18.60	4.62	70	.85	.72	.70	.72
	5	29	19.62	4.38	75	.85	.62	.62	.63
LITERATURE	1	22	11.81	3.43	53	.70	.56	.59	.57
	2	21	12.71	3.84	55	.77	.93	.86	.93
	3	21	11.33	4.00	49	.76	.44	.42	.44
	4	25	11.88	3.96	51	.78	.69	.66	.68
	5	25	13.48	4.05	58	.79	.54	.54	.54
HEART	1	23	17.93	3.92	70	.77	.86	.85	.86
COURTS	1	23	13.39	3.58	54	.65	.62	.63	.62
COMPUTER	1	23	15.08	5.97	60	.90	.76	.76	.76
ALL CTESTS (5 items)			74.28		60	.86*	.81	.81	.82
JUPITER	2	21	15.90	5.42	64	.71	.66	.63	.67
BLOOD	2	21	17.09	4.44	68	.78	.76	.69	.76
BACON	2	21	13.38	5.45	54	.78	.82	.76	.82
ALL CTESTS (5 items)	2	21	76.31		61	.90*	.86	.86	.88
ASTROLOGY	3	22	15.40	3.83	62	.73	.74	.83	.79
SOCIOLING	3	22	16.40	5.17	66	.86	.74	.76	.77
HORMONES	3	22	16.04	4.22	64	.76	.85	.77	.84
ALL CTESTS (5 items)			76.07		61	.91*	.83	.84	.85
PROBLEMS	4	25	17.49	4.04	70	.80	.77	.73	.76
CANALS	4	25	15.80	5.13	60	.87	.66	.62	.65
FRENCH	4	25	13.16	4.49	53	.88	.55	.46	.53
ALL CTESTS (5 items)			76.92		62	.88*	.81	.76	.80
WORK	5	29	19.82	4.81	79	.89	.72	.72	.71
CUISINE	5	29	18.66	4.03	75	.83	.70	.73	.91
FRESPEECH	5	29	15.68	4.66	63	.84	.61	.61	.62
ALL CTESTS (5 items)			87.16		69	.88*	.78	.75	.79
DICT		118	28.65	13.98					
GRA1		119	16.73	3.93	67				
VOC3		119	27.97	7.79	56				
GRA2		120	13.64	5.18	55				
GRA3		120	15.20	4.55	61				
VOC1		120	14.35	5.27	57				
JOINTC		118	31.09	7.79	62				

* calculated by Cronbach's Alpha
All others = KR-20

Table 2: Basic data English group

Test	Form	n	Mean	SD	P	r_{tt}
CARS	White	24	20.00	2.52	80	.52
CUISINE		24	19.08	2.13	76	.61
RICHARD		24	18.74	2.56	75	.59
CHIMPS		24	18.75	2.67	75	.55
ALL CTESTS (4 items)			76.52	7.34	77	.75*
PETS	Yellow	24	21.25	2.77	85	.72
AMERICAN		24	18.37	4.59	73	.73
CANAL		24	17.54	6.42	70	.93
CORNWALL		24	18.29	6.49	73	.93
ALL CTESTS (4 items)			75.45	17.16	75	.89*
TELEPHONE	Blue	24	19.38	5.56	78	.84
DIET		24	16.75	4.96	67	.89
WINDMILLS		24	16.08	5.04	64	.85
CONCRETE		14	13.20		53	.92
FIRST 3 CTESTS			52.21	11.65		.67*
ALL CTESTS (4 items)			65.41	17.54	65	.83*
BEHAVIOUR	Green	27	17.55	5.82	70	.89
RIVERS		27	17.81	5.10	71	.86
FLIGHT		27	15.25	6.29	61	.91
DECORATING		27	18.59	5.97	74	.91
ALL CTESTS (4 items)			69.20	21.04	69	.93*
				* calculated by Cronbach's Alpha. All others = KR-20		

are available. For the German subjects, who completed some of the same texts in the context of placement sessions at the University of Duisburg, extensive demographic and other data are available, including the results of the various subtests of the DELTA test (WS 90/91), a test whose own validity has been extensively investigated and been shown to be satisfactory (Lück and Klein-Braley, 1978, 1980; Schwibbe and Schwibbe, 1987). In Table 1 r_{tt} is the reliability coefficient, r_{DEL} is the correlation of the C-Tests and test parts with DELTA, r_{dict} the correlation with the dictation, and r_{TOT} the correlation with DELTA and DICT combined.

3.2 Text script analysis

A rich source of information is provided in the responses to C-Tests as revealed by the test papers themselves. The language processing techniques of individual test subjects and groups of subjects as demonstrated by their responses to C-Test items show quite clearly why C-Tests can legitimately claim to represent the concept of general language proficiency. I shall show how test-taking behaviour - as evidenced by the entries on the test scripts - enables us to make reasonable guesses about the nature of the language processing underlying the - correct or incorrect - responses.

3.3 Possible over-interpretation of data

The greatest danger inherent in an *ex post facto* analysis of test scripts is the possible over-interpretation of what one finds there. For instance, I contend that crossings out reveal reprocessing by the test subject and that incorrect responses can be interpreted as revealing specific psycholinguistic processes. I also contend that incorrect response behaviour is in many cases less trivial and careless than it seems.

To provide just one example here, given a context such as *Wine is generally cheaper than in a British hotel*, I shall argue that those subjects who originally wrote **home**, crossed it out and replaced it with **hotel**, engaged in reprocessing, probably as a result of their pragmatic knowledge that wine is not normally sold in homes, not even in British ones, despite the price of alcohol in England. Furthermore, I would claim that the reprocessing demonstrated here takes place at a very high level of text processing, the pragmatic level of virtually total text comprehension, so that although the change affects only one word, the difference in overall performance between those who left the word *home* standing and those who changed it to *hotel* is a meaningful one which can be interpreted in terms of the individual's standing on the language proficiency continuum.

3.4 Analytical procedures used in this study

For all the tests involved in this investigation normal statistical analysis on the super-item basis was performed; in addition, for every text an individual item analysis was calculated on the basis of the deletions in the texts. For some texts the responses of all respondents were transcribed, for others, notes were taken of particularly interesting phenomena. The selection of the depth of detail for the analysis of individual texts was determined by a variety of considerations, for instance, some texts were completed by more than one group, which made it possible to compare responses, others had been completed by very large groups of subjects, and so on.

It is not possible to do more in this publication than give a brief summary of the main findings; however, anyone interested in this approach will easily be able to obtain similar data, and will, I predict, quickly detect that the type of mechanisms I describe here operate in any set of C-Tests.

4. Response behaviour

There are three possible types of response behaviour to a C-Test blank. The examinee can give no answer at all, a wrong answer or a correct answer.

4.1 Unsolved blanks: early closure and narrow focus

There seems to be no way of interpreting unsolved blanks except as a breakdown in text processing. Statistically, we can compare the behaviour of all examinees with respect to the same blank and discover how difficult individual blanks are for the group overall, but a blank on its own cannot normally reveal why processing broke down at this particular point in the text. What *is* frequently associated with blanks is preceding *early closure*: the subject completes a sentence or part of a sentence in a way which seems syntactically and semantically correct at that point. However, because the unit has now been (apparently) completed, the subject cannot make any sense at all of the next blanks. Not all texts provoke or permit early closure.

Even more frequent is *narrow focus*, which, naturally, is also associated with blanks: often individual units are supplied for an item surrounded with blanks. Such items fit into the local context, but not into the macrocontext. Many examples could be given.

4.2 Different but correct responses

In some, generally in few, cases alternative responses can be viewed as correct. The number of permissible alternatives in a C-Test is usually small, but sometimes decisions about acceptability can be quite tricky.

4.3 Wrong answers

On the whole, wrong answers provide us with more insights into text processing strategies than right answers do. Respondents give a variety of different types of incorrect responses to the different items.

4.4 Incorrect spellings

One of the questions frequently asked by users of C-Tests is whether wrong spellings should be penalised in the same way as genuinely incorrect responses. From the point of view of test objectivity, it is preferable to count as correct only a response which either conforms to the original text or is accepted by native speakers as fitting into the text. Any scorer latitude opens the door to subjectivity. Statistically, it makes very little difference anyway (cf. Grotjahn *et al.*, 1992). But the question is, when *is* a word incorrectly spelled? Presumably ***hormon***, ***hormones*** and ***hormones'*** can all be interpreted as incorrect spelling of ***hormone***, but they can equally well be interpreted as genuine misapplication of rules. A large number of otherwise identical words need an additional ***e*** in English which is not present in German, so it could be claimed that those subjects who write ***hormon*** are still working within the German system. In ***hormones'*** an additional English rule (the greengrocer's apostrophe) has been misapplied, so this would surely be viewed as a mistake.

4.5 Singular / plural concordance

In a number of cases, the problem lies in the use of a plural where a singular is needed and vice versa. Otherwise the word is correct. This is, of course, one type of *narrow focus*. In some cases the focus is extremely narrow since the word which demands singular/plural agreement is directly adjacent to the damaged item. In other cases there are longer-ranging constraints: for instance, in the text HORMONES, ***variations*** is separated from the plural verb ***are***, which makes the plural obligatory, by six words. Interestingly, if the singular was offered in this item, it was repeated in the item five blanks (or ten words) later, where the same word was mutilated. I would hypothesise that the effort of keeping this repetition in mind over the length of the sentence was so great that accurate processing of the verb was impossible. This is an example of the phenomenon I call *system overload.*

This consistency of response over items replicates McKenna's (1978) cloze data: he found that students misinterpret later items in the test because of wrong solutions to earlier items. According to him, this phenomenon reveals that a representation of text meaning is built up gradually.

4.6 Right word class, wrong word

Some items show that examinees probably knew which type of word was required, but were not able to find the correct lexical item. Subjects show that they know the type of word needed, but cannot find the correct word. This happens with a large

variety of word types, including prepositions, infinitive forms, ***-ing*** forms, and so on. Other items show that subjects are trying to use their word-formation knowledge to supply the missing item. One German subject processing the text CANALS wrote above the blank ***lo__ 'Schleuse'***, indicating that she knew exactly what was needed, but simply did not have the requisite lexical item available in English.

Such responses offer evidence that the morphological, syntactical or collocational rule may be available, although the specific lexical filler cannot be found.

4.7 Near misses and nonsense words

Sometimes the subject almost knows the right word, but not quite. For instance, it seems to me that ***avariable*** (# 8, correct: ***average***) in HORMONES is an attempt to get at a word which is floating somewhere near the level of consciousness. This type of behaviour shows that the subjects have some idea of the word they want, but are not able to home in on it precisely. This ties in with the assumption that one of the ways in which entries in the mental lexicon are sorted is alphabetical.

Nonsense words are rare in C-Tests - examinees are much more likely simply to leave the blank unfilled, but they do occur. In HORMONES we find ***averalon*** (# 8, correct: ***average***) and ***smeed*** (# 19, correct: ***sex***). If nonsense words can be interpreted as a sign of the test-taker's indication to the tester that he or she feels unfairly treated, then the small number of cases occurring indicates that for examinees, at least, the face validity of the C-Test may not be as low as is often feared.

4.8 Suppressing / ignoring elements in the text

Not infrequently text processing appears to blank out portions of the text which interfere with the representation of meaning that the test subject is developing. One example is given in HORMONES # 7 where the solution supplied, ***control***, would make sense if the following word ***to*** is deleted, so that instead of ***differences in sex hormones would contribute to differences in behaviour*** the text would read ***would control differences in*** - a version which would also make sense, but which will not fit in the text as given.

4.9 "Careless mistakes" and system overload

I would attribute much apparently careless behaviour to the phenomenon already described which seems quite common in complex language tasks: the examinees who get this item wrong are suffering from *system overload*. The rule demanding agreement of subject and verb is not applied because although the subject *knows* the

rule, and would normally *use* it, at the actual moment of processing it is temporarily unavailable since too many other constraints are foregrounded. This is not the same problem as that of making the transition between declarative and procedural knowledge, because in other circumstances the rule would be correctly applied.

Theoretically, it seems plausible that system overload becomes less of a problem as the internal grammar becomes more efficient. The better students *do* suffer less frequently from apparently unexplainable howlers. The phenomenon as such is easiest to detect in translation into the foreign language where it can be shown quite clearly (Klein-Braley, unpublished research) that the same rule is correctly applied in other language activities such as writing essays. Translation can be viewed as more demanding since two languages are being processed, though not necessarily simultaneously. Completing a C-Test is also more demanding than simple reading or writing since both active and passive processes must be accessed to solve the items. Oral performance suffers from system overload too. Here, however, it is the result of the rapid rate with which the language elements have to be produced.

5. Detection of top-down processing strategies in action

It seems reasonable to assume that as language skills develop, so will the ability to make more use of the more efficient top-down processing strategies. One of the main problems with C-Test analysis is that although it seems logical that top-down processing must take place - because otherwise the text could not be satisfactorily reconstructed in its entirety - it is very difficult to detect such processing in action. This is partly a result of the very short texts involved.

5.1 Correct solutions and crossings out

Respondents who produce correct solutions to mutilated words are presumably producing their responses for the right reasons. Normally, we have no direct way of discovering what these are. However, in some few cases we can follow the examinee's thought processes to a limited extent, namely when someone changes his or her mind and crosses out the initial solution, replacing it with a second answer. Analysis of crossings out show that in the vast majority of cases the solution finally arrived at is the correct solution. In most other cases the tentative effort is crossed out, and no final insertion is made - the examinee gives up on this blank. The interim solutions finally rejected often reveal early closure or narrow focus, and rejection of the interim solution in favour of an unsolved blank shows that final decisions on acceptability/non-acceptability are probably made when the meaning of the text "as a whole" is available. Selection of a final correct alternative must almost certainly be made at a late stage in test processing.

5.2 Textual reconstruction: reinterpretation of the text

Textual reconstruction occurs when a test subject reinterprets the text so that it acquires a different meaning. Reinterpretation can operate on short- or long-range segments of the text. Occurrences of this are relatively infrequent, since not all texts lend themselves to reconstruction. Often, such reconstruction processes are the result of, or are accompanied by, early closure.

In HORMONES the following instances of reconstruction occur:

1. ***It has been wider assumed than*** *(it has been widely assumed that)*
2. ***It's an effort to prove this*** *(in an effort to prove that)*
3. *variation ... within each sex are tied to variation in behaviour (variations ... are tied to variations in behaviour)*
4. *with each sex are tied to the variable in behaviour*
5. *variations ... are tipical to variate behaviour*

The English children involved in the study engage in more reconstruction than the German university entrants. This could be interpreted as the result of having greater access to alternative units, or of less understanding of the task - not realising that they should recreate the original text as exactly as possible, which is probably a question of maturity. The most likely explanation in my view lies in a combination of these two things.

In general, such "revisions" of the text are unique to one person. Often, punctuation necessary for the new interpretation is inserted. Not infrequently, the text revisions extend over more than one deletion. Often the other deletions involved are processed correctly. In other cases the reinterpretation offered involves the sometimes explicit, sometimes implicit, deletion of a word, or part of a word, left standing, or, alternatively, the - often explicit - addition of elements not actually present.

Such reinterpretations are obviously the result of not entirely successful text processing, but they do show that longer-range constraints than merely the mutilated word and its immediate neighbours are involved. In SOCIOLINGUISTICS, for instance, we have the word ***social*** repeated three times, and each time the same respondents replace it with ***society*** and adjust the text as necessary, perhaps under the not unreasonable assumption that there could have been a mistake in the test construction process. It seems to me that such consistency represents the retention over a longer segment of text of the fact that this particular combination has already appeared once and been "satisfactorily" solved. This can be interpreted as showing the successful allocation and invocation of textual memory resources. This explanation is particularly plausible when the item repeated does not fit into the local macrocontext. When reprocessing takes place over a longer stretch of text with

invoked. Such response behaviour is an indicator that a successful - albeit erroneous - representation of meaning has been developed. It can therefore be viewed as evidence of high level text processing.

These data provide evidence that more able subjects do take longer-range textual constraints into account. The evidence can identify these processes only in those who make incorrect responses: there is no way to detect how those who complete the text correctly manage to do so. Neither the think-aloud techniques nor this type of analysis enable us to determine precisely the nature of the processing strategies used by the successful subjects, but it seems legitimate to assume that they are very similar to those whose solutions are incorrect.

6. Pragmatic text processing

The text CUISINE, an account for potential visitors of German and Austrian eating habits, enables us to look at two phenomena involving a link between the text and the outside world because of its cultural bias. The first one is the fascinating contrast between *home* and *hotel* in # 11: *wine is generally cheaper than in an English ho___*. A number of children who originally selected *home* subsequently crossed it out and wrote *hotel*. It is the higher scorers who tend to get this item right. Fewer of the German group (8 out of 29) chose *home*.

A second case where the differences in response between the English and the German group enable us to see knowledge of the world entering text processing is items 18 and 19 of CUISINE (***Germany and Austria***). None of the German group gets the concept wrong (although two students write ***Austrian***), but twelve (50%) of the English group write some form of ***Australia***. Obviously the Germans know that Germany and Austria are part of the German-speaking cultural landscape; English children associate ***Aus__*** with ***Australia***. In this case, there is no connection between getting the item right and making a better score. Presumably, the difference between the two cultural entities is that, in the first case, knowledge of the world has to be connected to logical probabilities in the text; in the second case, this is not important.

7. Conclusion

In this research I set out to discover what goes on linguistically when subjects complete C-Tests. Using the test scripts as sources, I was able to show that language operations must take place on many different levels in order to restore the mutilated text to its original form. The less proficient students work on the text at lower levels, the more proficient at higher levels. It seems, from the evidence, that subjects do not proceed through the text from blank to blank, but, in recursive text processing

cycles, gradually piece together successively larger sections of the text, reappraising earlier solutions in the light of the current macrocontext. To gain a score approaching 100% the subject must both understand the meaning of the text, have all the various items and structures available, and be able to conform with the necessary rules.

I believe that the evidence presented here shows that C-Tests can legitimately be viewed as tests of general language proficiency.

Bibliography

Feldmann, U., R. Grotjahn and B. Stemmer (1986), 'Was messen Sprachtests eigentlich? Überlegungen zur introspektiven Validierung von C-Tests', in Seminar für Sprachlehrforschung der Ruhr-Universität Bochum (ed.), *Probleme und Perspektiven der Sprachlehrforschung. Bochumer Beiträge zum Fremdsprachenunterricht in Forschung und Lehre*, Frankfurt am Main, Scriptor, 325-88.

Grotjahn, R. (1986), 'Test validation and cognitive psychology: some methodological considerations', *Language Testing*, 3, 159-85.

Grotjahn, R. (1987), 'Ist der C-Test ein Lesetest' in A. Addison and K. Vogel (eds.), *Lehren und Lernen von Fremdsprachen im Studium*, Bochum, AKS-Verlag, 230-48.

Grotjahn, R.(ed.) (1996), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*, vol. 3, Bochum, Brockmeyer.

Klein-Braley, C. (1996), 'Towards a theory of C-Test processing' in R. Grotjahn (ed.) (1996), 23-94.

Lück, H.E. and C. Klein-Braley (1978), 'Entwicklung des Duisburger Englisch-Leistungstests für Anglistikstudenten (DELTA)', *Berichte aus dem Arbeitsbereich Psychologie der Fernuniversität Hagen*, 2.

Lück, H.E. and C. Klein-Braley (1980), 'Der Duisburger Englisch-Leistungs-Test für Anglisten (DELTA)', *Diagnostika*, 26 (1), 93-4.

McKenna, M.C. (1978), *Diagnostic uses of the cloze procedure. Selected articles on the teaching of reading*, Set B, No. 41, New York, Barnell Loft.

Schwibbe, G. and M. Schwibbe (1987), 'Validierungsuntersuchungen zum Duisburg English Language Test for Advanced Students (DELTA)' in R. Grotjahn, C. Klein-Braley and D. K. Stevenson (eds.), *Taking their measure. The validity and validation of language tests*, Bochum, Brockmeyer, 255-74.

Stemmer, B. (1991), *What's on a C-test taker's mind: mental processes in C-Test taking*, Bochum, Brockmeyer., Bochum.

Stemmer, B. (1992), 'An alternative approach to C-Test validation', in R. Grotjahn (ed.), *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*, vol. 1, Bochum, Brockmeyer. 97-144.

Nine: The C-Test and L2 lexical acquisition/ processing research

David Singleton and Emer Singleton

Lexis has in the last ten years or so appeared to loom fairly large in publications aimed at the language teaching market (see, e.g., Carter, 1987; Carter and McCarthy, 1988; Gairns and Redman, 1986; McCarthy, 1990; Nation, 1990; Wallace, 1982; Willis, 1990). Interest in this area has indeed been such that after a considerable period in the wilderness, the teaching of L2 vocabulary has at last been restored to the agenda of teacher trainers and textbook writers. Unfortunately, despite this very welcome re-acknowledgment of the importance of the lexicon, it remains the case, as Meara eloquently demonstrated in his keynote address to the 1993 AILA Congress (Meara, 1993), that the state of basic research into L2 lexical acquisition and processing still leaves much to be desired. We believe that the C-Test has possibilities as a research tool which can play a part in remedying this situation, and in the present paper we attempt to justify that belief.

The paper begins with a first pass at the question of what elements of language are plausibly to be seen as included within the ambit of the lexicon. It goes on to discuss the aptness of the C-Test as instrument for eliciting data within the context of L2 acquisition and processing research, with particular reference to its capacity to tap into normal and "natural" language use. The focus of the paper then returns to lexical matters, detailed consideration being given to the issue of how the specific challenges posed by the C-Test relate to the current consensus regarding the nature of the lexicon. Finally, a brief account is given of the Trinity College Dublin Modern Languages Research Project and of the use of the C-Test within the context of this project.

1. Lexis and the lexicon: prolegomena

For many people language *is* (more or less) lexis and lexis *is* (more or less) language. A sense of this perception can be had from the ways in which language and language use are referred to in everyday parlance:

*I want a **word** with you.*
*Her **words** are perfectly clear on this point.*
*The **wording** is all wrong.*

Even the specialist study of language has had a somewhat "lexicocentric" history. In phonology, for instance, the test for phonemic distinctions using minimal pairs