

Longman Linguistics Library

**Elicitation Experiments
in English
Linguistic Studies in
Use and Attitude**

**Sidney Greenbaum
Randolph Quirk**



Elicitation Experiments in English

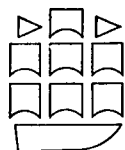
Linguistic Studies
in Use and Attitude

Sidney Greenbaum

*Associate Professor of English
University of Wisconsin-Milwaukee*

Randolph Quirk

*Quain Professor of English Language and Literature
University of London*



LONGMAN

LONGMAN GROUP LIMITED LONDON

Associated companies, branches and representatives throughout the world

© Longman Group Ltd 1970

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of the Copyright owner

First published 1970

ISBN 0 582 52452 0

Made and printed in Great Britain by
William Clowes and Sons Limited, London and Beccles

Preface

The experimental work underlying the present study was conducted within the Survey of English Usage and we are grateful for the ready co-operation of Survey colleagues, past and present. Our chief debt is to Jan Svartvik, who has given freely of his time and unique experience throughout: ideas on design, help in administering batteries, and detailed critical comments on a first draft of this book. Caroline Bott has been responsible for all the analyses of a statistical nature and for the programs by which the results of Batteries III and IIIa were computed. On the computational side more generally, we are happy to express our gratitude also to the staff of the Computer Centre of University College London; to Gordon Robbie of Her Majesty's Stationery Office; and to Ruth Kempson, who helped in this as well as in other respects. Valerie Adams, Margot Charlton, Derek Davy, Norman Fairclough, and Joan Huddleston also provided valuable help at various stages of the work, and in connexion with Battery IV we are glad to acknowledge the help also of Hannah Steinberg. The work was financially supported in part from a grant to the Survey of English Usage by HM Department of Education and Science and by a research grant for the summer of 1969 from the University of Wisconsin-Milwaukee. Finally, we are grateful to Peggy Drinkwater for her skill and friendly help in guiding the book through the press.

February 1970

S G

R Q

Contents

1 The aims of elicitation experiments	1
2 Experimental design	8
3 The relevant aspects of responses	19
4 Testing the test design	26
5 Comparability and consistency	37
6 The influence of experimental environment	50
7 Linguistic problems and scoring criteria	59
8 Linguistic problems and experimental variation	67
9 Use and attitude : the relation between test results	81
10 Conclusion	113
<i>Tabular appendices</i>	119
<i>Bibliographical references</i>	153

Figures

1	<i>Use and attitude</i>	1
2	<i>Types of test</i>	3
3	<i>Specimen preference test</i>	16
4	<i>Scoring system</i>	19
5	<i>Test and group variability</i>	41
6	<i>Consistency of individuals</i>	44
7	<i>Consistency of individuals</i>	46
8	<i>Consistency of individuals</i>	47
9	<i>Consistency of individuals</i>	48
10	<i>Attitudes to experiment</i>	51
11	<i>Attitudes to experiment</i>	56
12a	<i>Compliance and evaluation scores: category order</i>	84
12b	<i>Compliance and evaluation scores: category order</i>	85
13	<i>The relation of Figs 12 to Figs 14</i>	87
14a	<i>Compliance and evaluation scores: result profiles</i>	88
14b	<i>Compliance and evaluation scores: result profiles</i>	89
14c	<i>Compliance and evaluation scores: result profiles</i>	90
14d	<i>Compliance and evaluation scores: result profiles</i>	91
15	<i>Similarity and evaluation scores: result profiles</i>	104
16	<i>Selection and preference scores: result profiles</i>	108

Tables

1 <i>Compliance tests: Battery I</i>	120
2 <i>Compliance tests: Battery II</i>	122
3 <i>Compliance tests: Battery III</i>	124
4 <i>Compliance tests: Battery IIIa</i>	126
5 <i>Selection tests: Battery II</i>	128
6 <i>Selection tests: Battery III</i>	128
7 <i>Selection tests: Battery IIIa</i>	129
8 <i>Evaluation tests: Battery I</i>	129
9 <i>Evaluation tests: Battery II</i>	132
10 <i>Evaluation tests: Battery III</i>	134
11 <i>Similarity tests: Battery I</i>	135
12 <i>Similarity tests: Battery II</i>	137
13 <i>Similarity tests: Battery III</i>	138
14 <i>Preference tests: Battery III</i>	139
15 <i>Compliance tests: Battery IV</i>	140
16 <i>Evaluation tests: Battery IV</i>	141
17 <i>Group results - Compliance and evaluation tests: Battery II</i>	143
18 <i>Group results - Compliance and evaluation tests: Battery III</i>	146
19 <i>Group results - Similarity tests: Battery II</i>	148
20 <i>Group results - Similarity tests: Battery III</i>	148
21 <i>Group results - Selection tests: Battery II</i>	149
22 <i>Group results - Selection tests: Battery III</i>	149
23 <i>Group results - Preference tests: Battery III</i>	151

Chapter 1

The aims of elicitation experiments

The central concern of this book is the pursuit of experimental methods whereby grammatical and semantic inquiry can be put on a satisfying objective basis. It is true, but misleadingly curt, to say that our aims are to find out whether a given form is acceptable. Not merely is it misleading in as much as it implies a 'yes or no' decision: it is misleading also in implying that 'acceptability' is a simple, unified phenomenon. The techniques explored in this monograph are designed to cope with a multifaceted acceptability within which it is essential as a minimum to make the distinctions displayed in *Fig 1*.

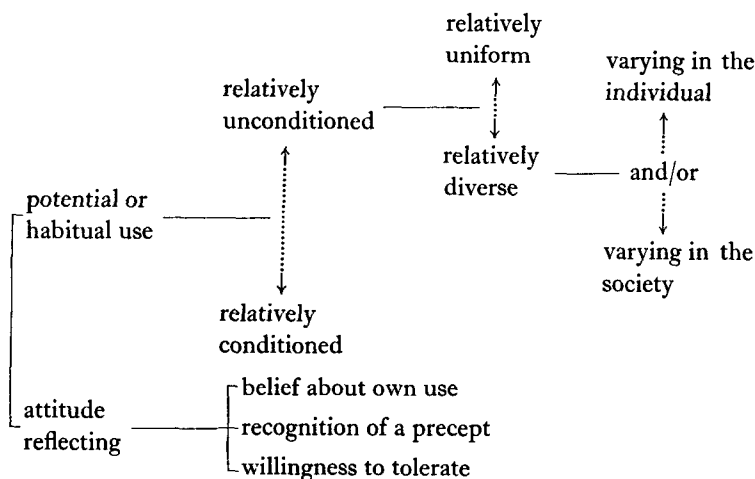


Fig 1: Use and attitude

It will be noted that 'actual' use is not mentioned in the figure; indeed, the purpose of our experiments is to move beyond the instances of actual usage (as recorded in a corpus) to the material for which elicitation techniques are required. But even so, we need a distinction between 'potential' and 'habitual' for material so elicited. That is to say, we must postulate the case in which the elicited sentence embodies essential features which a subject has encountered before (the past of *learn* or the occurrence of *hardly* between auxiliary and nonfinite verb, for example); this is what we understand by 'habitual'. But equally we must postulate the case in which the elicited sentence has an essential feature which a subject may never have been called upon to use but which is in some sense 'available' to him (within his 'competence', to change the metalanguage); this is what we understand by 'potential'. For example, we may wonder what a subject would supply as the past of a strange verb such as /flaiv/ or the way in which he would position the adverb *introductorily* or indeed (cf Jacobson 1964, 238) *blondely*. This is far from suggesting, however, that instances can be unambiguously tagged as 'habitual' or 'potential'.

A word or two may be added in explanation of some other distinctions made in Fig 1. The pair 'conditioned' and 'unconditioned' should be seen as polar terms on a graded scale, and the former should be read as 'conditioned by specifiable linguistic or situational factors'. A similar scale is of course indicated by 'relatively uniform' and 'relatively diverse', the latter pole being often referred to as 'free variation', a term avoided here because of its doubtful implications: one may question whether diversity is ever entirely unconditioned. Such variation may be a property of the individual, as when a Mr X vacillates between /sai'kɔlədʒi/ and /psai'kɔlədʒi/ in his pronunciation without this reflecting a similar vacillation in society as a whole. Equally, society as a whole may show variation between /i:ðə/ and /aiðə/ without this being reflected in the pronunciation of Mr Y, who says only /aiðə/. But these are not of course mutually exclusive and indeed it may be supposed that a variation in society usually corresponds to a comparable variation in the individuals who comprise that society.

If elicited behaviour is different from the 'actual' behaviour casually observed and (if one is lucky) collected in a corpus, it is at least equally important to distinguish elicited usage from attitudes to usage. And these attitudes can be seen as reflecting three potentially quite distinct but often interacting factors. We may have strong beliefs about the forms we habitually use and we may also have strong views about the forms that ought to be used; these may be in harmony or in rueful conflict, but –

needless to say – our beliefs about our own usage in no way necessarily correspond to the facts of our actual usage. Furthermore, we may tolerate usage in others that corresponds neither to the forms we believe we use ourselves nor to the forms that we believe are the most to be commended.

The discussion of our aims will be simplified if we outline at this point the experimental structure within which our inquiries have taken place. In *Fig 2* we display the types of test and the relation between them. It is important to emphasise the basic division between performance tests and judgment tests, since from each subject and on each problem we normally elicit responses in two complementary tests, a Performance test (our chief method of eliciting a subject's *use*) and a Judgment test (our chief method of eliciting his *attitude*).

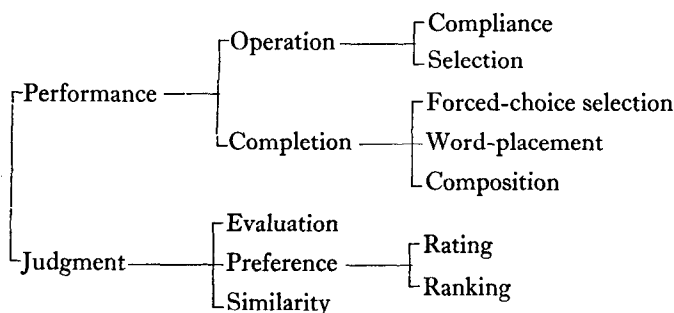


Fig 2: Types of test

Performance tests comprise operation tests and completion tests, differentiated according to the type of tasks required of subjects. For operation tests, subjects are asked to effect some change in a given sentence, while for completion tests they are asked to make some addition to a given sentence. Chapter 7 lists the categories of linguistic problem investigated in the operation tests we discuss in this book.

Operation tests comprise compliance and selection tests, which differ in the nature of the linguistic problem they pose. In compliance tests, some deviance is suspected in the sentence presented to subjects or in the sentence resulting from the change they are asked to make, except that some sentences thought to be non-deviant are interpolated as a control and for contrast. For example, subjects were asked to replace *he* by *they* in the sentence *He hardly could sit still*. It was predicted that the position of *hardly* in this sentence would be unacceptable to many subjects.

Indeed, by transposing *hardly* in their responses to the position between auxiliary and verb, the majority of subjects not only pinpointed what they found objectionable in the sentence, but also indicated by their alteration of the sentence what they considered to be the acceptable position of *hardly* in this type of structure. It was therefore not surprising that when these subjects were asked to perform the same task on the sentence *He could hardly sit still*, they retained the adverb in this position. In some compliance tests, however, the deviance emerges only as a result of the task required of the subjects. Thus, subjects were asked to turn into a question the sentence *He will probably stay late*. There is no reason, of course, to think that the sentence given to them was in any way deviant. But the sentence resulting from the application of the task raises the problem of the acceptability of *probably* in the interrogative form of the sentence, the extent of the problem being indicated by the alterations that subjects carry out.

Selection tests are devised as a method for investigating divided usage. They are rather like the last type of compliance test in that the sentence presented to the subjects is not thought to be deviant. However, when subjects perform the specified task, they have to choose (whether or not they are conscious of choosing) between two or more variant forms. For example, subjects were asked to make the verb present in *None of the children answered the question*. The task obliged the subjects to choose between the singular and plural forms of the verb and their responses automatically indicated which form they preferred to use.

Completion tests comprise forced-choice selection tests, word-placement tests, and composition tests. The first of these, like the ordinary selection tests described above, are devised for the study of divided usage. The difference is that subjects are given a limited set of items from which to select and a limited set of environments in which their selected form is to be used. For example, given

I ----- the poem.

I have ----- the poem.

together with the two forms *learned* and *learnt*, we investigate not what is the subject's general preference as between *-ed* and *-t* (as in ordinary selection tests) but his preference for one or other form as preterite or past participle if he is forced to choose. See further Quirk 1970b and Kempson and Quirk 1970.

As the term indicates, word-placement tests are designed to investigate word position. Subjects are given a sentence and a word that they are

required to use with it. For example, subjects were asked to use *usually* with *My brother plays the guitar*. By their placement of *usually* they indicated which position in the structure they preferred for this adverb.

Composition tests, unlike the other completion tests, are open-ended. Subjects are given part of a sentence and are told its position in the putative final form of the sentence. They are then instructed to complete the sentence in any way they like. For example, subjects were given *I entirely* as the opening words of the sentence they were required to complete. In this particular instance, we were interested in the verbs which co-occur with these opening words as compared with those co-occurring with, for example, *I completely*. An analysis of the results of composition tests conducted so far as well as a discussion of further uses of this technique will be found in Greenbaum 1970.

Three types of judgment tests have been used: evaluation, preference, and similarity tests. In evaluation tests, normally complementary to compliance tests, subjects are required to evaluate a sentence on a three-point scale: 'perfectly natural and normal', 'wholly unnatural and abnormal' and 'somewhere between'. For example, subjects have been asked to judge in this way the acceptability of a sentence they had been given earlier in a compliance test, *He hardly could sit still*.

Preference tests, on the other hand, are normally complementary to selection tests. They comprise two components, rating and ranking. Subjects are given two or more variant forms of a sentence, for example *None of the children answers the question* and *None of the children answer the question*. They are then required to rate the sentences, using the same three-point scale as in the evaluation test, but this time the juxtaposition of the two forms inevitably focuses evaluative attention on the only variation between them. They are also required to rank the sentences in order of preference.

The similarity test also involves a judgment on the relationship between sentences, but this time it is the semantic relationship that is in question. Subjects are given two sentences, usually with minimal lexical and syntactic differences between them, and are asked to judge their similarity on a three-point scale: 'very similar in meaning', 'very different in meaning', and 'somewhere between'. This judgment was required, for example, on the pair of sentences (given here in a prosodic transcription explained in Crystal and Quirk 1964)

/some lectures are actually given before tèn#

/actually# /some lectures are given before tèn#

Similarity tests, including this particular instance, have generally been used as complementary to compliance tests.

Not all these types of test have been used to an equal extent up to the present, and indeed in this book we are concerned chiefly with batteries that have consisted predominantly of *compliance* tests with their corresponding *evaluation* tests. The results in each battery are displayed in the Tabular Appendices.

As already stated, our techniques aim at exploring all these facets of acceptability, but this is not a matter of whether a form is acceptable or not. Acceptability is gradable and we are concerned with the extent to which a sentence is unacceptable. But of course we also want to know the precise point at which it is unacceptable, and in what respect it is unacceptable. The latter, for example, is in important ways deducible from the direction that 'rectification' of a deviant sentence takes at the hands of subjects in a compliance test. More positively, sufficiently varied types of information are sought so that we can establish the normal position of adverbs in declarative and interrogative sentences, the normal process of negation, and such like. For example, with a test sentence *The council lowered his rent slightly* and the requirement that the subjects make the verb present tense, more explicit information on the subjects' normal usage is given by those who move *slightly* than by those who leave it unchanged. Moreover, among those who found it acceptable in the evaluation test there were some who moved the adverb in the compliance test; such a discrepancy illustrates the contrast between an *attitude* of tolerance and a preference in *use*. It is also important to recognise that a preference in use need not correspond to a preference in attitude. For example, with the subject phrase *None of the children*, a singular verb (prescribed by schoolroom precept) was preferred in a judgment test more frequently than it was preferred in a selection test. On the other hand, we need to recognise that there is in general a fairly close harmony between attitude and use, as is shown in Chapter 9. For example, about a third of the subjects gave as their first choice in the preference test *I have smelled the flowers* and about a third also supplied the form *smelled* in the selection test where the same sentence was involved.

In this connexion, we may mention that the preference tests give us an obvious example of the way in which our work can distinguish between variation in the individual and variation in society (*Fig 1*). They show that some subjects can prefer one form while other subjects prefer a different one (variation in society); and they also show that some subjects give an identically high ranking to two or more forms (variation in the individual).

It is hoped that it will now be clear that our scheme of test techniques (*Fig 2*) can be regarded as a promising source of data corresponding to the categories of use and attitude (*Fig 1*). Some of these categories of course respond more satisfactorily and sensitively to experimental inquiry than others, and there can be little doubt that 'habitual behaviour' is the most difficult to ascertain by such means. We do not claim that a simple equation can be made between this and even the most overwhelming results our experiments can elicit. We are well aware that we cannot escape from the artificiality of the test situation, though with continuing refinement we can hope to remove some of the worst effects of bias that the test situation introduces. Meantime, we can be sure that the alternatives are considerably less promising: reliance upon corpus alone and reliance upon introspection alone. Both need supplementation by experimental evidence.

Chapter 2

Experimental design

Our experiments are an extension of the techniques developed in Quirk and Svartvik 1966 (hereafter QS). In QS, a battery consisted of 50 performance and 50 judgment tests, allowing the whole battery to be administered well within a lecture period. We have seen no reason for any general departure from this convenient battery-length. The QS tests were characteristically in pairs which contrasted deviance and non-deviance in respect of a single linguistic feature. As well as such pairs, there were non-deviant sentences where the tester's interest lay in the specific form selected by subjects. As a result, about half the test sentences were non-deviant and this was (and has continued to be) regarded as important in order to avoid the development of an expectation of deviance.

The tests were administered orally by the tester standing in front of a group of subjects, the only control on consistency being the fact that he read from a prepared script which specified the forms of instructions to a prearranged plan and which prosodically specified the way in which the test sentences should be read out. The instructions, which gave no hint that the sentences to be heard would include deviant ones, listed the tasks and gave examples of their performance. The test sentences came at timed intervals, the intervals (of 20 seconds' duration) being measured from the beginning of one test to the beginning of the next. Tasks stipulated for operation tests were as follows (QS 22f): change of tense to present or to past; change to negative or to affirmative; replacement of a subject pronoun by a stated singular or plural form; turning a statement into an 'inversion question' introduced by a specified form of *be* or *do*. These tasks were introduced in a varied order, but 'paired' tests (though always widely separated in the test sequence) required the same task. In

blocks of roughly ten, the procedure varied as to whether subjects were told first the task to be performed or the sentence to which a task was to be applied.

Until the instructions for the subsequent judgment tests (in all cases, of the evaluation type), the tester was careful to mask from the subjects the fact that the experiment had anything to do with linguistic acceptability. But at the beginning of these instructions, subjects were told they would hear 'the same sentences again, this time at much shorter intervals', and they were now invited to judge their acceptability on a three-point scale:

Wholly natural and normal

Marginal or dubious

Wholly unnatural and abnormal

For the evaluation component, the sentences read out were the same as the sentences given in the compliance tests (before task performance) and they were presented in the same order. The sentences came at timed intervals (of five seconds' duration), again measured from the beginning of one test to the beginning of the next.

Apart from batteries not primarily designed for eliciting linguistic information, batteries since QS have been far more homogeneous in respect of the linguistic features being tested. We are here primarily concerned with those hereafter referred to as Batteries I, II and III. For the most part they were designed to investigate certain aspects of adverbial use and Battery I dealt exclusively with such problems, paying close attention to the multiple use of an adverb (as disjunct, conjunct, or adjunct, to adopt the distinction in Greenbaum 1969a). For this purpose, certain common adverbs were used several times with varying degrees of differing function; in 16 sentences, two adverbs were used in a somewhat similar sentence frame and in eight of them the same adverb appeared twice in order to test the extent of contrast recognised by subjects. For example,

1. /hōnestly# /Mr Jones honestly reported our stōry#
2. /rēally# the /students réally wōrk during the term#

Furthermore, in attempting to achieve maximum comparability in results, contrasting adverb uses and positions were tested in identical or near-identical lexical environments. For example, 'His sons completely managed the family business' (the first sentence in the battery) and 'His sons managed the family business completely' (the 26th sentence).

The test was successful to a great extent in that subjects introduced fewer changes in sentences (1) and (2) above than in (3), where the two adverbs are in a more tautologous relation:

3. /nevertheless# /some people nevertheless attempted it#

Thus, one of the instances of *nevertheless* was omitted by 11 out of 85 subjects, whereas with *honestly* and *really* omissions were made by only two and three subjects respectively. An attempt had been made to counteract any tendency to monotony by changing the frequency with which we switched from announcing the task first to announcing the sentence first; instead of keeping the procedure constant for blocks of ten, we narrowed the variation down to blocks of two. Useful in itself, this innovation could not of course offset the repetitiveness of linguistic pattern, which tended to invalidate the results. For example, in several instances where a sentence had two different adverbs (as in 'Frankly, the workers were honestly answered by the manager'), subjects responded with a sentence which had two instances of the same adverb ('Frankly . . . frankly . . .') instead. Moreover, so habituated did subjects become to having two adverbs in sentences that in three cases where they were given sentences having only one adverb, they added a second one. In one case, in fact, this addition resulted in two uses of the same adverb, so that 'His sons managed the family business completely' became 'His sons completely managed the family business completely', though this may result from the earlier occurrence of the sentence, in which *completely* was preverbal.

It was therefore decided to re-introduce greater variety into later batteries, especially so far as sentence pattern, lexical variety and type of sentence deviance are concerned. On our reading of the results of Battery III, it seems to us that a satisfactory balance has been struck between a degree of homogeneity that yields useful quantities of comparable linguistic data, and on the other hand a variation of sentence content and pattern that prevents subjects from having the feeling that there is repetition. The nearest we came to pattern similarity was with the following three sentences, which were widely scattered throughout the battery and which were accompanied by only one or two other sets showing similarity:

you could /always send it this afternoon#
 you should /always take it before dinner#
 they could /always go there tomorrow#

The re-ordering experiments (see below, p 32) gave us no reason to suspect that memory of earlier instances in the above cases had a vitiating effect on the results.