LINGUISTICS

TREES OF DESCENT IN LINGUISTICS
AND TEXTUAL CRITICISM

ARCHIBALD A. HILL
The University of Texas

There has been much discussion of late years over the processes by which language-forms, and consequently languages, change. The central issue is probably whether or not change occurs as a modification of the speaker's rules of production, a question which (unfortunately) I can not approach. However, there is a sister science which has close connections with linguistics, at least in its historical aspects — the study of the transmission of manuscript texts. Some of the problems of the two sciences are close enough, so that it seems worthwhile to return to a consideration of them for the light such a consideration may throw on the edges of the area of current discussion[1]. The connections between the two sciences are in two areas. The first is simply the area of spelling, where linguists have not always been as cautious or as knowledgeable as they might be. The second is the area of differentiation, where in expressing language relationships and copying of texts, the two sciences have made use of family trees.

As an expression of the current scepticism about matters of language change once regarded as firmly settled, I was more than a little amused at a statement made at a recent meeting of the Linguistic Society of America. The statement was that all that we really 'know' about sound change is that forms were once pronounced in one way, and are now pronounced in another. The statement struck me as amusing, and as an instance of insufficient awareness of textual scholarship. I should rather say that all that we really 'know' about anguage change is that forms were once spelled in one way, and are now spelled n another. Everything that goes beyond the primary data — the spellings—

---

whether it is a matter of guessing at pronunciation, of guessing at a manner of change, or of guessing at forms not preserved in spelling, is inference. Inference, of course, is form and structure in our science, but not its foundation. It is therefore, doubly important for all of us to know the nature of our foundations in data before we attempt to infer from them.

If I say that linguists are sometimes less than clear about the nature of their data, I can perhaps do so without sting by using one of my own mistakes. I found, in a popular account of the Battle of the Cowpens, a form spelled *Thickelly Creek*. This was Tarleton's spelling of the form *Thicketty*. I promptly evolved the theory that Tarleton had heard an American voiced, flapped /t/ — the sound I used in *water*. It would have been reasonable to suppose that such a sound was strange to him, and that he had interpreted it as the nearest sound in his own speech: — a voiced, flapped, lateral. Unfortunately for what still strikes me as a pretty theory, a friend pointed out to me that Tarleton had used a map in which ⟨t⟩s were regularly uncrossed.

More importantly, because the data and the conclusions are more closely related to general historical questions in linguistics, we have not always paid enough attention to distribution of spellings in classes of forms. In the handbooks still in use, it is stated that Old English had a sound [æə] which developed out of an earlier [æ] before [l] plus consonant. Such a word as *healf*, with the spelling ⟨ea⟩ is a typical example of the change, called breaking. A generation ago it would have been stated unhesitatingly that the new form developed in only one environment, and also never failed to develop in that environment. As a result, the new form [æə] was believed to be incapable of contrasting with the old form, [æ], since the contrast necessarily rested in the environment, and the vocalic difference was therefore believed to be redundant. But in standard West Saxon there are a few forms in which the spelling ⟨æ⟩ is regular, under conditions similar to those in *healf*. One such form is the adjective *ælmihtig*[2]. I would once have said without hesitation that such a form proved the breakdown of complimentary distribution, and that the two sounds were — after the adoption of such spellings as that of the adjective — in contrast. Now I would assume that the fixing of the spelling in a few forms was the result of the adoption of a spelling difference, itself caused by pronunciation development, as a means of distinguishing the identity of the meaningful forms, much as we differentiate the spelling of *quean* and *queen* or *coarse* and *course*. Since the spellings ⟨ea⟩, ⟨a⟩, and ⟨æ⟩ are distributed as unpredictable variants in most forms, it seems to me that the later conclusion I outlined is the

better of the two, though there is a third possibility, that the old form of pronunciation was preserved in this form in a fashion causally related to its identity. This last view is now fashionable, though I dislike it since it seems to me to minimize structural regularity. I suggest, therefore, that spelling evidence rules out one of three conclusions, and leaves the second and third undecided. I do not need to labour my point — whatever conclusion a linguist settles upon, he must examine the spellings of his manuscripts first. We as linguists are required, therefore, to know something of the techniques and conclusions of analytical bibliography.

It is a commonplace that a historical scholar must somehow establish his text before he describes its linguistic characteristics. In all instances where he finds himself dealing with a text preserved in multiple copies, his principal tool has been the tree of descent, drawn up according to the methods of Karl Lachmann, the founder of 19th century manuscript scholarship. In order to express the relative authority of multiple copies of a text, Lachmannian scholars draw up diagrams, or trees, showing the closeness of each copy to the author's original. The trees, therefore, are a means of reconstructing the text as it came from the author's hands, free of corruption.

The most basic of Lachmann's postulates is that any two manuscripts which have readings in common, must have a common source. To this postulate there can be no objection. Lachman insisted further, however, on the separation of readings into right readings and errors. The right readings belonged in the author's original, and the errors were the result of later corruptions.

Unfortunately, the division of readings into right and wrong, leads to circularity. For one thing, any reading which is structurally possible, and which occurs in an older text, is likely to be right, even if totally unsupported, and unknown in meaning. To state that a structurally possible reading is an error involves the assumption that we know all about the language of the past, when in fact we know very little about it, and can never know it all. The history of manuscript scholarship is dotted with brilliant emendations which later research has proved to be unnecessary. As for the circularity, it is worth pointing out that to divide readings into right and wrong is to make conclusions about relative authority before the study has been begun. That is, the editor first selects right readings, then uses them to establish authority, and then uses authority to select the right readings that he prints. In such circumstances, one wonders why an editor bothers to draw up a tree at all; he might equally well rely merely on his literary taste, as Bedier suggested that he should do[3]. For those who wish to investigate the attempts of recent manuscript scholarship to free itself

[1] Stanley, E. G. 1969. "Spellings of the Walden group" in *Studies in language, literature, and culture of the Middle Ages and later*, ed. by E. Bagby Atwood and Archibald A. Hill, University of Texas at Austin: p. 67 *ælmihtig* was counted 66 times, all with ⟨æ⟩, by Stanley. The closely similar *alwalda* was counted 15 times, 13 with ⟨a⟩ and 2 with ⟨ea⟩.

[3] I have given my views on the use of 'right' and 'wrong' readings as the basis for manuscript trees at length in "Some postulates for distributional study for texts", *Studies in bibliography*, University of Virginia, 3 (1950-51), 63-95. The views of Bedier are discussed on pages 86-90.

of the dangerous reliance on right and wrong readings, I can suggest the reading of Greg's *Calculus of Variants* [4].

A second misleading Lachmannian assumption has been the belief that relative authority is directly related to the date of copying. We all are acquainted with editions of older texts in which the editor bases his work on the oldest manuscript. Yet since the drawing of a manuscript tree expresses only the number of intermediate versions which can be proved to have intervened between two copies of a text, data of copying is irrelevant. A copy made in the 20th century may have been taken directly from the author's original, while one made in the 15th century may have been at the end of a long line of intermediaries.

Perhaps the best way to show the importance of considerations of this sort to historical linguistics is to take up one matter which at one time was almost proverbial — the 'final e's in Chaucer'. The rules believed to govern the occurrence of final ⟨e⟩ in Chaucer were worked out by scholars like ten Brink and Kittredge, who were strongly influenced by neo-grammarian principles, then the most vital force in linguistics. Their endeavour was to establish the hypothesis that final ⟨e⟩ occurred in Chaucer's lost original in strict accordance with etymology. Thus all etymological final ⟨e⟩s were right, all others wrong. The rightness of final ⟨e⟩s was then used to establish the relative authority of the surviving texts, and since all texts had passed through the hands of scribes, editors felt free to bolster their conclusions by rejecting all unetymological ⟨e⟩s. The result, not unnaturally, was to present the etymological rules as fully unassailable. What is needed, therefore, is surely a re-examination of texts to see exactly what the manuscripts say, and then in return, a rewriting of the rules.

Some idea, indeed, of the size of the task that still, after so much time, awaits the historical students of English can be gained if I point out that I once surveyed the editions of the Early English Text Society for the purpose of evaluating the accuracy of texts. Most texts seem to have been edited with no very firm principles of method of any sort, and even the relatively few critical texts were primarily Lachmannian. Even such a monumental work as the Manly and Rickert *Chaucer* often presented an unnecessarily complex picture of the descent of manuscripts on account of the editor's desire to establish authority for all readings literally excellent enough to be worth preserving. Thus then, historical linguists need to know enough of manuscript method so that they will no longer be at the mercy of editors. Historical linguistics and bibliography should, indeed go hand in hand. Yet in order to be strictly fair, it should be pointed out that bibliographers should know more than they often do about linguistics. Even such a bibliographical giant as Greg attempts to define classes of variants

[4] Greg, W. W. 1927. *The calculus of variants*, Oxford: Clarendon Press.

on the very rough basis of whether a variant does or does not affect meaning, without having first come to a conclusion about the nature of meaning.

I have been maintaining that historical linguists need to know something of the nature of manuscript work, so that they can be more wisely critical of editorial policy. Trees of descent occur, however, in statements of linguistic relationships, as I indicated earlier. I believe furthermore, that the trees drawn up by linguists for strictly linguistic purposes may have been unduly influenced by manuscript trees because of lack of understanding of the nature of the data that each class of trees was meant to express.

I have pointed out earlier that date of copying is irrelevant in manuscript work [5]. Thus there is nothing inherently illogical in a tree which shows manuscripts A through F all at exactly the same distance from the common ancestor X, since such a diagram does not mean that all the manuscripts were copied at once. The copying of the various manuscripts might perfectly easily have been, in fact probably was, separated by long periods of time.

In the descent of languages, and language-forms dates of differentiation and change are always relevant. To cite a commonplace example, the three sets of changes involved in Grimm's 'law', Verner's 'law', and the Germanic stress shift can only have occurred in that order, unless our whole system of inference is wrong. That is, Verner's 'law' operated on spirants produced by Grimm's 'law' in accord with a stress distribution older than that produced by the stress shift. The result is clearly that language communities differentiate, just as forms do, in chronologically successive stages. Thus the High German Sound shift, giving *Wasser* and *Fuß* against English *water* and *foot* is later than the Germanic changes which produced the voiceless stops still found in English and Scandinavian languages.

The implications of these truisms have not always been realized by workers in historical linguistics. One of the still greatly and rightly respected handbooks in historical study is Bright's *Anglo-Saxon Reader*. In a printing of this book as late as 1935 there is a diagram of considerable naïveté at page xviii. Aside from the curious fact that Bohemian, Polish and Lithuanian are represented as the three independent branches of Balto-Slavic, the diagram gives eight daughter languages of Indo-European, all directly descended from the parent language in a single step [6]. The picture seems to me almost laughably improb-
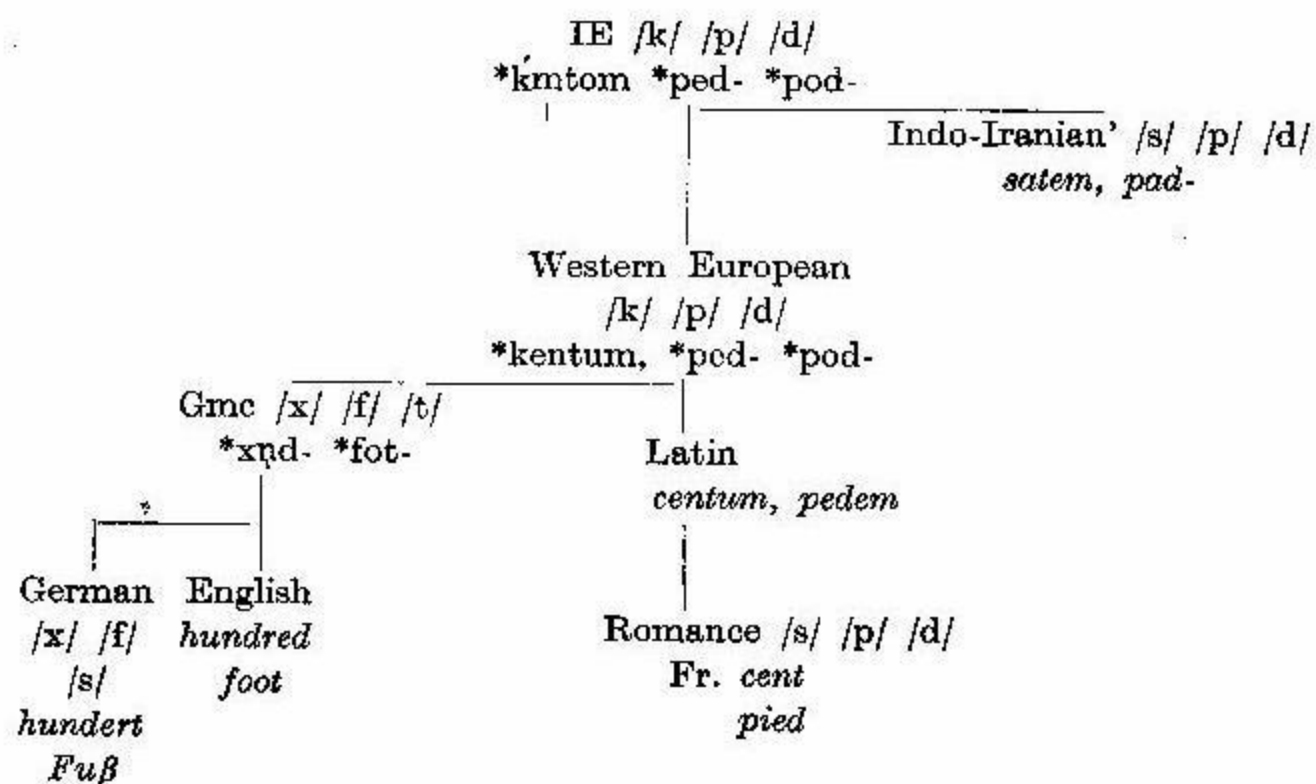
[5] An additional reason for regarding the date of copying as irrelevant in drawing up a manuscript tree is that all families of extent manuscripts must be supposed to have been decimated, so that readings preserved only in a late copy might be shown to be ancestral to those preserved in a much earlier copy. An example of such an authoritative, but late copy is discussed in "Postulates for distributional study", p. 85–86.

[6] Bright's *Anglo-Saxon Reader*, revised by James R. Hulbert, New York: Holt and Co., 1935, p. xviii. The diagram has never, to my knowledge, been corrected, though the *Reader* was reprinted in 1959.

able — as if there had been a sort of first International Congress of Indo-Europeanists at which it was decided to set up eight independent languages at four that afternoon.

Not only do language forms evolve through time, it seems to me also probable that they evolve by successive bifurcation. Again to use commonplace examples, the Indo-European /k/ of the word for *hundred* split first into the /s/ of Iranian *satem*, and the Latin *centum*, preserving in that language a form more nearly like the original than that found in Iranian. The Western /k/ type developed only later into the voiceless velar spirant of Germanic, giving a new pair /k/ : /x/, to put beside the older /k/ : /s/. The three types /k/ : /s/ : /x/ did not develop all at once. And as if to clinch the matter, the /s/ found in French and Spanish forms of the word for *hundred* did not develop at the same time as the Iranian /s/ of *satem*, but only much later, out of Vulgar Latin *centum*.

To establish that languages, as total entities, develop by similar successive bifurcations, it is necessary to suppose that linguistic change is systematic, in that a single process affects a multiplicity of forms, so that a large number of changed items can be referred to a single change. This is the fact back of general statements of change like Grimm's law. That is, the /ð/s of English *father, mother, brother* all show the same correspondence to Latin /t/s as in *pater, mater, frater*. (There are details of history that can be conveniently forgotten in this discussion, however.) Further, it is reasonable to suppose that some of these general developments are sufficiently pervasive in the structure of language to justify us in setting them up as the dividing points at which the language communities in question were differentiated. A proper tree for the series of consonant changes we have been discussing would then look like this:

IE /k/ /p/ /d/
*k̑mtom *ped- *pod-

Indo-Iranian' /s/ /p/ /d/
satem, pad-

Western European /k/ /p/ /d/
*kentum, *ped- *pod-

Gmc /x/ /f/ /t/
*xṇd- *fot-

Latin
centum, pedem

German /x/ /f/ /s/
hundert Fuß

English
hundred foot

Romance /s/ /p/ /d/
Fr. *cent pied*

The table above differs from that of a typical family of manuscripts in more ways than just that each joint is a binary split. It is also assumed that it is normal for a change to consist of the splitting off a divergent form, and the continuation, in a straight line, of a more conservative form. Such an assumption is not thought of as a universal rule, however. There are instances of change in which both parties to the split are modified, and indeed, modified in such a way as to make it impossible to say that one form is divergent, the other conservative. In fact, the split into *centum-satem* involves a minor change in the *centum* half, since the IE ancestral form is believed to have been a palatal voiceless stop, whereas the *centum* group treat the descendant as merely a voiceless dorsal stop without reference to palatal or velar quality. In this instance, the descendants do not prevent us from describing the western half of the family as the more conservative. With the IE velar stops, however, the situation is more complex. The forms found in such words as Gk. *gynē*, (woman) English *queen* never appear without modification in any daughter language, making it much more difficult to set up a divergent and conservative pair of developments.

The difference between conservative and divergent developments is shown on the tree by the representation of the conservative developments in a straight line, and the divergent elements as departures at right angles. Incidentally, a development which results in no splitting, is merely shown by a straight line with representation only of the early and late forms, as with Latin *centum*, followed later by Romance *cent, ciento*. Finally, in the tree given, the relative spatial closeness to an ancestral form shows the relative date of a divergence. Thus the development of the French *cent* is much later than both the Gmc. development of */xṇd/* and Iranian *satem*.

A final word is that in instances where it is not possible to show that one form is conservative, the other divergent, the relationship can be conveniently shown by diagonal lines. This is the case of Old English 'long ā', as in *hām* which has developed into a front nucleus in modern Scottish /heym/ and a back nucleus in Modern Standard English, *home*. A word can also be said about the assumption that splitting is by successive bifurcation. I said earlier that this kind of splitting was that which I assumed to be normal. The limiting case may be instructive. A good many years ago Douglas Chretien, in oral discussion, called my attention to the differentiation of the Polynesian languages, which occurred in a very short space of time by colonization of widely separated islands. Though in point of fact the various colonial ventures did not occur exactly simultaneously, they did occur so close together, and the contacts thereafter were so sparse, as to make it more realistic to view the differentiation as multiple and simultaneous than as successive binary splitting.

Up to this point we have been considering mostly ways in which manuscript

and linguistic trees differ. Our next point is one in which they are similar. Trees in manuscript and language show conflicting results for differing items. In a manuscript tree it is often the case that the majority of readings show that manuscript x has been copied from manuscript y, but that there are a few readings in which the copyist has consulted the manuscript z, of quite a different place in the total tree; and in these the scribe has taken his readings from the second source. In manuscript work, such conflicting sources for readings are shown by drawing dotted lines to show the source of the conflicting readings, and the process of drawing readings from elsewhere than the main exemplar is given the name 'contamination'.

A similar situation in language descent can be shown by the same example we have been using throughout. The Slavic sub-family within IE is a member of the *satem* branch, Gmc of the *centum*. Yet the two sub-families share a host of characteristics, particularly in etymology, not found elsewhere in IE community. Often taken as an example of the special closeness is the presence of an otherwise unknown dative plural in -m, clearly the result of Slavic-Germanic innovation. To represent such a contradiction on a family-tree diagram, it would be necessary to show the two twigs, Slavic, and Germanic, each springing from a separate main branch, as growing together again. In linguistics, of course, this state of affairs led to the famous 'Wellentheorie', in which the relationships are shown as a series of overlapping circles[7]; Germanic is outside the *satem* circle, but Germanic and Slavic are both inside the circle for dative plural in -m[8].

It is now generally recognized that both types of language diagrams are necessary, and that conflict in the spreading circles is the result of the fact that contact between peoples of similar, but not identical speech, goes on after some differentiation has taken place. Similarly, scribally mediated contact between differing exemplars can take place in manuscript descent. There is, indeed, no reason why the types of diagrams used in either type of work, could not be employed in the other. In fact, in still another type of work, the mapping out of semantic differentiation and development, a similar state of affairs exists. Fodor and Katz, in a well-known discussion of the word *bachelor*, present the several meanings of that word as the result of a series of binary splits, capable of being represented on a family-tree diagram. Yet some of the meanings seem to conflict with such a branching representa-

tion. Once again a dotted line representation, or a wave-theory diagram could be used to resolve the difficulty. And once more the difficulty arises because the separate entities on the diagram are in contact with each other through the mediation of the language-user[9].

The implications of conflicting differentiations in manuscript and language work seem fairly clear. They imply that binary differentiation is seldom the whole story. In short, linguists and bibliographers must both allow for cross--class development, and remember that such development is an important limitation on the pervasiveness of binary differentiation.

Yet in another direction, particularly in the reconstruction of starred forms, it seems to me that linguists have not gone enough in recognition of the normality of binary differentiation. All too often linguists have a tendency to rely naïvely on a simple majority for recognition of the older form, and then to use this majority form as the basis for reconstruction. If the principle of binary differentiation is accepted, it is no longer possible to reconstruct older forms by mere counting. There are instances, it is true, in which the majority form is a guide, but these are limited. They are those instances where (other than the development on which the tree is based) the one half of the tree shows unity, while the other half shows spotty diversity. Most of the time, however, we find that we arrive at a point in which there are only two forms, an A and a B, each representing a single branch on our tree. Such a state of affairs is normal, if language develops by binary splitting. Curiously enough, if we look at manuscript tree as drawn by editors of older texts, we also find that binary trees, giving only two main branches, are also common—I think, indeed, too common. That is, if an editor draws his tree so as to show only two equal branches, he leaves himself free to choose the readings, which appeal most to his taste, and thus escapes from the 'tyranny' of a triply branching tree, in which he is forced to take the readings of two against the third throughout. I have argued elsewhere that binary trees in manuscript work are suspicious for this reason, and that a conscientious editor ought to accept such binary trees only with the utmost caution[10]. It is here necessary only to add that in language work, binary splitting is the usual state of affairs, but that in manuscript descent multiple copying from a single exemplar is the normal process.

Why have linguists been slow to adopt a binary approach to language descent? I think it is primarily because they were under a somewhat uncritically adopted influence from Lachmann. That is, linguists would have been

---

[7] The overlapping circle or Wellentheorie diagram is most readily available in Leonard Bloomfield's classic work, *Language*, New York: Henry Holt, 1933, p. 316.

[8] The use of overlapping circles is, however, in accord with the results of geographical differentiation of dialects, where overlapping circles, 'isoglosses', are much the most convenient way of showing that, say, the development of the vowels in the Dutch words for *mouse* and *house* do not have neatly coinciding distributions. See Bloomfield, p. 328.

[9] Katz, J. J. and J. A. Fodor. 1963. "The structure of a semantic theory", *Language* 39. 190. *Bachelor*, "young human, male, unmarried", is at the opposite side of the tree from *bachelor*, "young, animal, male, unmated seal". It would seem that the two uses are closer than the first is to "one who has the lowest academic degree", or "knight serving under another knight".

better off with either less influence from Lachmann, or under an influence which adopted only those things which can be shown to be reasonable in linguistics, rejecting those reasonable in manuscript work but unreasonable in linguistics, and even more importantly, those things unreasonable in linguistics because they were also unreasonable in language work. No less a student than Bloomfield himself seems to me to have been open to the criticism which I have just given. On page 310 of *Language* he says

"The comparative method assumes that each branch of language bears *independent witness* [italics mine] to the form of the parent language, and that identities or correspondences among the related languages reveal features of the parent speech".

The phrase 'independent witness' is a common bit of manuscript-scholar's jargon, and is evidence enough of the Lachmannian influence. It is also true, that Bloomfield would seem to be thinking of reconstructing hypothetical forms much as a manuscript scholar reconstructs the forms of the author's original.

The interesting fact is that in the next pages, Bloomfield goes on to say (quite rightly) that languages have never been completely and simply unified or differentiated — that there never was a single parent *satem* dialect. The discussion goes on to the presentation of the overlapping circle diagram I have already mentioned.

I find nothing to object to in what he has said — only to what he did not go on to say. In linguistic history, the differentiation of single items — whether we study the pronunciation of *water* and *butter* in the Germanic languages or the syntax of the copulative verb in American dialects, the development of each separate item is, I believe, always the result of binary splitting, though unfortunately for students, always with only partial coincidence with the splitting in other items. In manuscript work, the differentiation of readings is again always binary in any given act of copying an individual item, but always with results only partially systematic and consistent with the differentiation of other items. My quarrel with Bloomfield is that the binary quality of individual developments is lost sight of, in the necessity of systematizing inconsistent individual histories. Perhaps the device of coining a slogan might help to remind us of the importance and the limitations of binary differentiation. A linguistic law of differentiation can be said to occur when a maximum number of items show the same development.

[10] "Postulates for distributional study", p. 90.