DOCUMENT RESUME

ED 049 467 AL 002 794

AUTHOR Mackey, William Francis
TITLE Interlingual Distance.
INSTITUTION Laval Univ., Quebec. International Center on Bilingualism.
PUB DATE 71
NOTE 139p.; Presented at "The Quantification of Multidimensional Differences" at the Conference on Contrastive Linguistics and Language Universals at the University of Hawaii, January 11-16, 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS *Bilingualism, Computational Linguistics, Connected Discourse, *Contrastive Linguistics, English, French, *Language Universals, Semantics, Taxonomy
IDENTIFIERS *Interlingual Distance

ABSTRACT

The measurement of interlingual distance (how far removed one language is from another) is both possible and feasible; and it can be computed in different ways. The difference between the codes of the two languages can be measured by one technique and the differences in samples of discourse by another. The samples may be measured as static entities covering a certain space, or as dynamic or kinetic sequences unrolling in time. The distance between two languages may be measured as the sum of their differences or as the amount of work necessary to convert one language into another. Each can be measured either by taking all the characteristics in which two languages can differ and counting the number of differences in each, or by integrating the immediately observable differences into a single formula of measurement. (This study of interlingual distance begins with a discussion of language universals and types of interlingual distance, followed by a discussion of semantic and formal differences and taxonomic and integral distances in discourse. Figures and tables, using French and English as contrastive illustrations, are included.) (Author/AMM)

INTERLINGUAL DISTANCE

by

William Francis Mackey

International Center for Research on Bilingualism

Laval University, Quebec

1971

Prepared for the Conference on CONTRASTIVE LINGUISTICS AND LANGUAGE UNIVERSALS at the University of Hawaii (11-16 January 1971) and presented as : The Quantification of Multidimensional Differences

INTERLINGUAL DISTANCE

Contents

INTRODUCTION

It has been said that all languages are dialects. Not all dialects, however, can afford to be languages, unless their speakers are independent enough to say they are. The mutual incomprehensibility of German dialects like Plattdeutsch and Bavarian has not been sufficient to raise them to the status of languages, no more than the inter-intelligibility of their northern cousins has reduced the Scandinavian languages to the status of dialects. If mutual intelligibility is a criterion for distinguishing language from dialect, at what point of similarity does the one change into the other? We do not even have the measurements needed for finding the point. And if we did, we could ask better questions, such as, how far removed is the one from the other. This is just one of the needs for developing measures of interlingual distance. There are several others in such fields as language learning, language teaching, language contact, bilingualism, creole studies, and in the geographical and genetic studies of languages.

In the field of language learning, it has been commonly accepted that some languages are more difficult than others, especially when seen from the viewpoint of the second language learner.

It may be observed, simply by looking at a text of any two languages that there are differences, not only in the way they look and sound, but also in the number of elements we may recognize from knowledge of our own language. Once we start trying to learn these languages, however, the differences become much more apparent. The number of new ways of pronouncing or writing will vary according to the language chosen; so will the number of differences in such things as word order, agreement, and length of the equivalent elements in our own language.

But how much more difficult one language is from another is a matter of conjecture and argument. Some estimates have been made by observing how long it takes certain persons to reach a certain level of proficiency as measured by certain tests. For example, according to a study made for the United States Peace Corps, it takes almost twice as long to learn Chinese as it does Malay; or more exactly, if it takes an educated adult native speaker of English 720 hours to achieve a certain level of proficiency in Malay, it will take him 1320 hours to attain the same level of proficiency in Chinese.

In the teaching of foreign languages, it has always been difficult to determine the extent to which elements of the mother tongue may profitably be transferred to usable habits in the foreign language. What is the degree at which such interlingual identification becomes useful?

In the field of bilingual education, ethnic minorities are often labeled by the language they speak - Spanish, German, French, and so on - and it is this language that is sometimes decreed as being their normal educational vehicle. In fact, however, the variety of the language actually understood by the minority may be an admixture of regional dialect and foreign words far removed from the international standard that educators have in mind. At what point are such dialects worth salvaging as media of instruction? In other words, if instruction is to be given in a language other than that of the

country, is the local variety of that language near enough to an educated standard to make its use worth while? Or is it so far removed that it will become more of a hindrance than a help? Again we do not have a way of finding out the degree of difference to enable us seriously to investigate the problem.

As for the bilingual parents, to what extent is their speech influenced by one language more than the other? And in the case of actual language mixture and creolization, to what extent is each language related to the parent tongue? What sort of mixture is it? And what are the proportions of the mix?

Translation studies, interpretation research and the field of comparative stylistics could eventually benefit from a quantification of their methods of confronting two different languages.

In the study of language relationships, it is useful to discover the extent to which geographically adjacent or genetically related languages have come together or diverged; but we would need standard measures of the extent of such similarity or difference.

It will be seen later that many of these questions are related - that, for example, interlingual intelligibility is a function of congruity, which in turn is a function of bilingual interference, and that both are related to the process of interlingual distance. Indeed, research into all these areas could benefit from accurate measurement and comparability. It is for the purpose of advancing the design of such research that the elaboration of techniques of interlingual measurement was undertaken.

The first problem is to find out the extent to which one language or dialect can differ from another. This will depend on our knowledge of how languages differ and of what they have in common, that is, on the universals of language. It also will depend on what we mean by interlingual distance and what we consider to be representative measurement. Our immediate concern will, therefore, be with universals, types of distance and techniques of measurement. Let us first examine the question of universals.

1

LANGUAGE UNIVERSALS

1.1 Man: Mind, Body and Environment

1.2 Language: Code, Discourse and Communication

1.3 Universals of Human Speech

1.3.1 Characteristics and Constraints

1.3.2 Common Components

1.4 Universals in Interlingual Distance

1. LANGUAGE UNIVERSALS

A central problem in the study of language universals is to find out how languages do not and cannot differ. To find out how languages do not differ would require a systematic examination of all the languages of the world. To find out, on the other hand, how languages cannot differ would require an inquiry into the nature of man in relation to language and the nature of language in relation to man.

Since we are not yet prepared to examine systematically all the languages of the world, we will have to forego our examination of what all languages do have in common and concentrate on the question of what all languages must have in common. We can attempt this by postulating the characteristics which all human beings must share and what all languages as languages must comprise. We can then study the relationship between these universals and make use of the results to distinguish different types of language distance and possible ways of measuring them. Let us first consider the postulates on man and his language.

1.1 Man: Mind, Body and Environment

It would seem plausible to assume that all human languages have common human traits. If all men are essentially the same in body and mind, are there certain physical and mental characteristics related to language which can be said to be universal?

We can perhaps make the following general assumptions about man and his speech:

1. He can hear the sounds he uses and make the sounds he hears.
2. He can put sound groups together into texts.
3. He can direct these texts to different hearers and modify them to suit different hearers.

4. He can associate the sounds he hears and learns with perceptible differences in his environment.
5. These environmental and sound features he groups into functional units.
6. The grouping is done for him by the language he hears, and they become conventions to which he is bound.
7. He also learns to associate analogous grouping with the non-environmental workings of his mind - with logical and conceptual processes.
8. In this way, all man's mental activities become associated with these conventional groupings of sounds and concepts - his perception of the sound sequences and what they stand for, his conception of them into equivalent categories of thought, his cognition of the relationship between them, his evaluation of the speech effects upon persons he hears and addresses, his use of these relations and evaluations to find solutions to his questions, his memory of all this, and his use of what is thus stored in anticipation of future thought and expression.

If certain elements of man's mind and body are indeed the same whoever his is or wherever he may be, then certain association of his mental activities in language with his physical actions in speech may well be universal. There are, for example, certain limits imposed by the make-up of man's thought and speech; he must function in a time sequence and his speech must operate within the limited capacities of the human anatomy - excluding as they do, for example, such anatomically impossible sounds and apico-pharyngals.

If, in addition to this, all environments in which man is able to exist

have something in common, some of these features may also be universal. Man, in order simply to exist must eat, drink, sleep, act and reproduce. And we can consequently suppose that all his languages must have labels or concept categories which include these essential activities. We can therefore say that some characteristics of man's mind, body and environment are universal by their very nature. We can also assume some universal characteristics which all languages by their very nature must share.

1.2 Language: Code, Discourse and Communication

All languages must include certain code, discourse and communication components. The codes of all languages must be composed of symbols and of procedures (or rules) for putting them together to generate messages (texts or discourse) usable for purposes of communication. All codes have certain essential elements in common. Although their elements must all be finite, their generating capacity may be infinite. Their elements are not things in themselves; they are representatives (or symbols) of things. These symbols may represent universal operations of the mind, like identity (is) and negation (not), universal features of the human condition (experience and action), universal features of message production (reference to something said), and universal features of all actions of communication (person speaking and person spoken to). Each code contains the symbols it needs to operate. All codes need symbols for (1) logical and human operations, (2) discourse features and (3) communication variables. Each symbol may have one or many forms expressed in the elements which all language have in common. Each type of symbol may be related to the essentials of man's mind, body and environment. The possibilities of these relationships may be seen by plotting one set against the other. (See Table 1)

TABLE 1
AREAS WHERE UNIVERSALS MAY BE SOUGHT

Language \ Man	MIND	BODY	ENVIRONMENT
CODE	1	2	3
DISCOURSE	4	5	6
COMMUNICATION	7	8	9

In each of these nine areas, we may ask which features are universal. What are the universal characteristics of 1. transformations (propositions,etc.), 2. articulation (phonemes, word forms,etc.), 3. categorization (size, number,etc.), 4. utterence (identification, substitution, reference,voice,etc.), 5.catenation (assimilation, distribution, etc.), 6. collocation (compounding, word linking, etc.), 7. style (mood, forms of address, etc.), 8. prosodic expression (intonation, tempo, etc.), 9. register (formality, slang, etc.), media (spoken, written).

1.3 Universals of Human Speech

What is universal about man's impressions and the way he labels them to produce speech? Is there something common in the physical sources of these impressions, or the process of applying labels to them? Are there common components applicable to all human speech? All human speech may have something in common in the processes which produce the characteristics and constraints of any language and the components common to all of them.

1.3.1 Characteristics and Constraints

It would seem evident that there must be symbols in the minds of men to enable them to talk about what all men have in common - their bodies and their

needs - eating, drinking, walking and sleeping. Although these symbols may well be universal, their characteristics are not necessarily so.

Although words for parts of the body, for example, may well be universal, there is no universal agreement on what can be considered as a part of the body. Some of the Slavonic and Celtic languages, for example, label the foot and the leg as one part - Serbian noga; Erse cos. They may likewise combine arm and hand - Russian ruka. Some languages label the extremeties with the same word - Spanish dedo for both toe and finger. Like other languages, however, they can specify by using phrases or compounds. In Spanish, they can say, los dedos de los pies. In Korean, however, they say, palkarak (toe) and sonkarak (finger). And in this language, as in others, some parts of the body may be extended or be extensions of a more general category; as in the above example, toes, fingers, spoons and other instruments can be labeled as a certain kind of stick (karak): palkarak (toe), sonkarak (finger), sukkarak (spoon), chotkarak (chopstick). This may seem unusual only if it is someone else's custom; English, French and German and other European languages have similar extensions. In English, there are the hands of the clock, the face of the clock, the foot of the table, the eye of a needle, where French uses needle (aiguille) of a clock instead of hand. But like English, in other contexts French makes a distinction between needle and pin (épingle), whereas German has a more inclusive nadel label for both, using compounds to distinguish between needle (Nähnadel) and pin (Stechnadel).

Many but not all of these metaphorical processes touch the grammar of languages and they may be formalized in any of the grammatical features such as position, word-ending or formal marking, like affixation. The distinctions may be logically formalized but arbitrarily applied; for not all languages apply logical categories in the same way. We may think, for example, that the distinction between person and thing is universal; but we need only compare the

grammars of English and French to find that such is not the case. French makes no formal distinction between person and thing, either in its pronoun system (il, elle) or in its articles (le, la, un, une); English does (he, she, it), but not always, not in its articles (the, a). The universals of logic are not the universals of grammar.

We must distinguish between what a language has and how a language functions. Most language seem to have logical categories; but they are not always logical in their use. A concept categorized as a thing in one language, may be classed as an action in another language. Time in one language may be treated as space in another (A long road takes a long time to cover.) Because of this arbitrariness of languages in the use of universal logical categories, it would be hazardous to guess which grammatical features are common to all languages.

Difference in categorization appears as a constraint on the conventions of the language. It is as if those who speak it have agreed to a package deal which they must take or leave. Such constraints are typical of all human organization. If a pharmaceutical chemist wanted to open shops in France, Germany, Canada and the United States, in France and Germany he could limit himself to the sale of medicaments; in the United States and Canada he would also have to deal in cosmetics, magazines and a host of other products. Contrariwise, if his business man friend wanted to sell soap in these countries at the retail level, he would have to open a perfume store (parfumerie) in France, a non-prescription Drogerie in Germany and a pharmacy or drugstore in the United States and Canada. Or if a man were interested in telegrams, he would be dealing with the post office in England, France and Germany, with the railway companies in Canada, and with a telegraphy company in the United States. But if he were involved in inter-urban telephone communications, although he would also be dealing with the post-office in England, France and Germany, he would have to associate with utility

companies in the United States and parts of Canada. In a similar way, the package deal of language categories functions as a binding feature - not only semantic, but formal as well. It also extends down into the neurological mechanisms of speech, operating as associated nerve reflexes. Native speakers of English, for example, when they utter vowels involving action of the back of the tongue are subject, through mutually conditioned reflexes, to the constraint of lip rounding at the same time. The reflex prevents them from disassociating the one from the other, making it difficult to pronounce foreign vowel sounds where the one is not bound to the other, as for example, in making the proper distinction between such French words as rue and roue (see below). Bound reflexes are the manifestations, at all levels, of the way a particular language groups the features of sound and meaning into the elements of its code.

Even though all languages may have labels for nouns, verbs and qualities, they do differ in the features of the environment which they can categorize under each of these labels, that is, in what is considered to be a thing (noun), an action (verb), or a quality (adjective and adverb). This "as if" process touches all levels of language, while, from language to language, it is differently applied.

It may be that the more abstract or general we get when talking about language, the more common characteristics we are likely to find. All languages have sentences and can express propositions. All languages label things. The way these labels are put together in propositions may have been determined at the dawn of human speech by the fundamental dichotomies of man as he evolved into a tool user. It has been suggested that when man succeeded in establishing a dichotomy between the working or hammering hand and the passive or holding hand, he paved the way for other dichotomies in human thought and speech. Did this also enable him to make an analogous dichotomy in the types of impressions that reached his

senses and the grouping of them into the stable or holding ones (the names, the subject), and the hammering or operational ones (the actions, the predicate), and to generalize this, in cognition, into all topics on the one hand and all comments about them on the other? It has been observed that the distinction between actor and action in child utterances occurs about the time the child has learned to make a distinction between one hand and the other. There seems to be a parallel between the child's physical development and its linguistic development. Systematically pairing physical and language observations in the development of children of all languages might therefore throw some light on the existence of language universals, and on the common components of code, discourse and communication.

1.3.2 Common Components

Are there components which all language codes have in common? Are there also common components in the discourse and communication aspects of language? To what extent are language codes alike? Are there latent systems and structures in the very makeup of the human brain which are transformed by the growing child into the system of a particular language? If there is a universal grammar, does it have a biological counterpart? It seems that the nature of the mind's symbolic behavior may be universal; in language it stores and operates a system of symbols so organized as to permit productivity and efficiency in communication.

Although these symbols are related to the environment which the code of the language divides into concept categories, some of the symbols are so general as to relate to features common to all environments, some of them probably inate to the very workings of the human mind - cases like the impossibility of a thing being and not being at the same time and under the same circumstances, negation, distinctions between things and qualities, actions and states, place and time, actor and object, container and contained and the numerous possible

relations between things and people. If all languages, for example, distinguish between actor and action, or subject and predicate, can we further assume that they distinguish between who-subjects and what-subjects, between modifications in time (when) and space (where), between the why and the how of an operation? Do they also all distinguish between being, having and doing? If they do, we could also argue for the existence of universal core sentence types expressing such relations as: Who is who, What is what, Who did what, Where is what, When is what, Who has what, Who does what, etc. If, in addition to this, all languages can distinguish between the causes of an action and the results, we could assume that they also express such relationships as "Who did what to whom, Who did what how, etc." If this is true, certain case concepts like the accusative, dative, locative, causative and temporal, no matter how expressed, may also be universal.

We can perhaps assume that all languages have grammatical means for distinguishing actions from the causes of such actions and their effects. That is, they have verb-like categories and noun-like categories, and also perhaps quality-like categories. But how far can we go in these speculations on the universals of language codes?

Can we say as much about the common components of discourse? There may be universals of discourse, such as the necessity of referring to something said before, and the need to substitute a shorter form for a cumbersome repetition. Here again languages differ in how they satisfy this universal need. French will put previous reference into the same category as its deictic (Cet homme...) in cases where English will consider it as definition (The man..). All languages have ways of putting together their concept and grammatical categories to yield messages - ways of connecting words in collocations (E. glass of milk/

G. *Tasse Milch*), conjunctions of words and phrases (*come and go*, *the sooner the better*), order of action and actor, predication and voice (*who did it*, *whom was it done to*), and of connecting statements or propositions in different types of relationship (*Although*, *Because*, *Since*, *While*, *If*.....*he did it*.). In other words, all languages must interconnect their elements by putting them together in meaningful discourse - which could be regarded as their sequential linking of concept categories. It is this interconnection of elements that enables language codes to operate in discourse with great economy - to have labels standing for many different things and filling different functions - elements which are in themselves inherently ambiguous until eliminated by verbal or situational controls. The efficiency of this is seen by the rarity of ambiguity and the few situations where the contexts themselves are ambiguous - and languages have alternate means of expression to clear up such ambiguities.

If the universal requirements of logic affect the discourse of languages, so do the universals of communication. In order to make language function as a vehicle of communication between two individuals, it may be argued that all languages must make distinctions between questions and non-questions, between the person spoken to and the person speaking, ways of expressing the attitude of the one toward the other. The attitude toward the person spoken to may be expressed by the type or length of the form, categorizing people in different ways (e.g., F. *tu*, *vous*/ G. *Du*, *Sie*). Relationship to the person spoken to may also be conveyed in the assumption and use of shared symbols as reflected in slang and technical jargon. Similar processes may be used in different languages to reflect the attitude of the speaker to what he is talking about. If the speaker is commenting on the fact that someone refuses to spend his savings to buy a new car, the choice of words like *thrift* or *avarice* reflects the speaker's attitude toward the person spoken about. Here again, the way these distinctions

are applied may be far from consistent. They languages use the same form for the person spoken about and the person spoken to - especially if it is a polite form. Languages also have ways of expressing what is expected of the person spoken to - (inversion of order may be used to illicit a statement from him (Are you?), a special word (Who.....?) or simply a change in the tone of voice (interrogative intonation).

What can one conclude from this incursion into the cloudy heights of speculative linguistics? It seems that languages may have a lot in common in what they can say, but that they may differ unexpectedly in the way they say it. Without actually analysing all languages, the only certain universals are likely to be found in the make-up of man - his body, particularly his speech organs and in the nature of the discourse he produces as a sequence of symbols in his necessary acts of communication. Of what use can these universals be in measuring the distance between languages?

1.4 Universals in Interlingual Distance

How can we use any of the universals of mind, body and environment and what is common in codes, discourse and communication to measure the differences between languages?

If we know what these universals are, we can use them as parameters. For instance, we know all organs of man used in speech and we know their physical limits. These can be used as constants in phonetic measurement. We also know certain things that all men have in common - time, space, parts of the body and we can use the number of distinctions made in each language to measure interlingual distance. Between these two extremes of language, which touch the physical world of sound and meaning lie a multitude of possible differences in systems and structures. The different grammatical organizations of language

can only be measured indirectly. And we will probably not be able to get exact measurements until we know what grammatical elements - like the parts of speech - happen to be universal. Not knowing this, we can either assume that there are no universals of grammar, or after all languages have been compared in like fashion, we can extract the constants found in the grammars of all languages.

Meanwhile, these constants - if there are any - of how languages can but do not differ may constitute a uniform error in all measurements of intergrammatical distance. For example, if both French and English have distinctions in their grammars between actor and action, this common feature, if taken into account, reduces the distance between the two grammars. If, however, this distinction turns out to be universal, the distinction has no longer any differential significance.

Certain universal characteristics of man, thought and communication, can therefore be used as measurements of interlingual distance. These can be broken down into distinctive features very much like a television signal is analyzed and synthesized to reproduce a picture on a screen. In this way, each phoneme can be broken down into distinctive sound features, just like each concept category can be considered as being made up of a number of semantic components.

Combinations, sequences and modification of these features may be permitted or obligatory, in both what must be expressed and the way in which it is. Combination of lip rounding and vowel fronting, as noted above, are permitted in French /y/, but not in English, where lip rounding makes tongue retraction obligatory /u/. Sequences of consonant and /-s/ are permitted in English (pillars, plurals); but not in Spanish (pilares, plurales). These constraints are part of the significant differences between languages and would have to be taken into account directly or indirectly in any complete measure of language distance.

In what way then do the universal characteristics of man and his language relate to the measurement of interlingual distance? The universal characteristics of man supply the measures; the universal components of language indicate the types of measurement. Since all languages must have some code, discourse and communication features, it is these that will have to be taken into account in divising types of measurement. The distance between the codes will not be the same as the distances in the discourse they produce; and interlingual distance in oral communication will not be the same as it is in written communication, since the static features on the printed page may differ in many ways from their kinetic counterparts in the stream of speech.

These universal distinctions of language relate in turn to the universal characteristics of man. The human body is the vehicle and producer of speech and as such can be used to measure the distance between speech codes. The universal characteristics of man's environment are relevant insofar as they are accounted for in the language codes. The universal characteristics of man's mind, its symbolic processes, its mechanisms for transforming code into discourse, and even its tolerance of inconsistency, will affect any attempt to measure the differences between languages.

Between these abstract universals of the human mind and the concrete universals of the human body, lie the multifarious diversities of the languages of the world. The universals enable us to study the diversities by revealing all possible types of interlingual distance and by serving as instruments for measuring them.

2.

TYPES OF INTERLINGUAL DISTANCE

2.1 Distance in Code and Discourse

2.2 Static Distance and Kinetic Distance

2.3 Distance as Difference and Conversion

2.3.1 Differential Distance

2.3.2 Conversion Distance

2.4 Taxonomic Distance and Integral Distance

2.5 Semantic Distance and Formal Distance

2.6 Direct Distance and Indirect Distance

2.7 Diversity, Intensity and Productivity

2. TYPES OF INTERLINGUAL DISTANCE

The universal distinctions between mind and body, code and discourse, speaking and writing, mother-tongue and other tongue and the diversity and intensity of man's conception of his environment oblige us to consider the measurement of interlingual distance from a number of different viewpoints.

These considerations make it evident that distances between languages can be measured from different angles and that there are different types of interlingual distance. We can distinguish between: i) distance in the two codes and distance in the discourse produced through them, ii) distance as communication of static messages in space, as distinct from different texts unrolling in time, iii) distances as differences between languages, as opposed to distance as movement from mother-tongue to other tongue, iv) distance as an analytic or taxonomic accumulation of categorical differences, or distance as an over-all integrated measure, v) distance as a direct measure of difference, in contradistinction to indirect measures making use of an intermediate language, vi) distance as the diversity, intensity and productivity of the comparable differences between the languages.

Since not all these different types are mutually exclusive, the distance between two languages may be measured from a number of different angles, as will be demonstrated below. But before elaborating techniques of measuring interlingual distance from these different angles, let us examine what the different types of distance involve.

2.1 Distance in Code and Discourse: (A X), $\binom{A}{X}$

As we have already seen, all languages have systems of symbols variously referred to as *langue* or code, and these are used in conventional ways to produce messages or texts - an activity which has been called by a number of names including *parole* and discourse. Measuring interlingual distance from the one is not likely to be the same as measuring from the other.

In measuring the distance between Language A and Language X therefore we can proceed either from the descriptions of the languages or from samples of them. In other words, we may measure the distance (D), (1) in the codes D(AX), or (2) in the messages, or samples of discourse $D\binom{A}{X}$.

Measurements of the differences between two language codes depends on the accuracy and comparability of the descriptions used. If the information supplied is either incorrect or incomplete, for one or both languages, the consequent comparison or contrast is likely to be inaccurate. If the type or methods of description (and notation) used are not identical, differences in descriptive techniques may appear as alleged differences in the languages. That is why some of the contrastive studies which are based on the two codes have their source not in the use of ready-made descriptions but on re-descriptions of the languages for purpose of comparison - or contrast. This is another reason why two contrastive descriptions of the same languages or dialects may differ.

In measuring from discourse samples, however, sets of equivalent texts with alternative versions are needed. But even with an inventory of all possible alternatives, it will be seen that measuring from the message is not the equivalent of measuring from the code. For one cannot assume that different codes will produce correspondingly different equivalent texts or that the degree of difference between two equivalent texts corresponds to degrees of difference in the codes which generate them. This is evident when one travels from one area to another area in which the same language is spoken. There are situations, for example, in which Englishmen say "Not at all", Canadians say "You're welcome," and Americans say "Uh, uh." Yet these three equivalent discourse responses to the same situation do not indicate that we have three different languages.

Between two different languages, the different codes may produce equivalent alternatives which are closer than others. For example, the English sentence The name is Robert may be rendered with equal correctness in French as either

Je m'appelle Robert or Le nom est Robert, the latter being much closer to the English equivalent. This does not mean, however, that the two French sentences are interchangeable; they are, on the contrary linked to mutually exclusive situations, the first being used in introduction, the second in consultation. The English sentence is used in both; to do so in French would consequently constitute a case of interference. Measuring language differences through equivalent texts therefore is valid only where both texts operate within the same context.

Similarities in code do not necessarily produce similarities in discourse. When measuring from the code, the English and German vocabularies may turn out to be very close, as it is in the German sentence: Kann (can) man (one) ein (one) Boot (boat) haben (have)? But measuring from the English equivalent of this sentence, the two languages no longer seem so similar:

G. Kann man ein Boot haben?/ E. May we rent a boat?

Although such code similarities may facilitate comprehension of the other language, they may be a source of interference in expression. The codes of the two languages may be close, while their respective transformations into equivalent messages may be highly divergent.

Contrariwise, the identity of two equivalent messages does not mean that the codes producing them are likewise identical. Sentences like Maria va a casa may be read as either Spanish or Italian. But if one were to compare the two codes which produced this same sentence, the two languages would appear as different: Sp. (pres. ind.): voy, vas, va, vamos, vais, van./ Ital. (pres. ind.) : vado, vai, va, andiamo, andate, vanno.

Spanish	Italian
Yo voy a casa.	Io vado a casa.
Maria va a casa.	
Nosotros vamos a casa.	Noi andiamo a casa.
Teresa y Maria van a casa.	Teresa e Maria vanno a casa.

The distance between two languages, therefore, is unlikely to be the same when measured from their codes as when measured from their messages. We must conclude that neither measure is alone sufficient to give the complete picture; we require both. And since they are so very different, we must consider measuring separately, distance between codes and distances in discourse.

2.2 Static Distance and Kinetic Distance: D, D

Distance between languages will also be different depending on whether we measure the static results of language communication or the dynamic process of the generation of texts from their respective codes.

In static measurement (D), language is considered as if it were a reservoir of material to be matched and measured. Two languages may be treated as if they were two landscapes, each corresponding feature of which is given equal consideration. Language can also be regarded as a kinetic process (D), the dynamic generation of texts on an ever-ending time scale. And the point on the scale at which elements of languages are equivalent can be regarded as a function of their difference.

Since discourse is essentially sequential, it can be regarded as unrolling in time, so that at no given time is a whole text perceptible. In this way, the resemblance of what comes before makes us anticipate what may come after. And because of this, the matching elements that come first have more resemblance value than those that appear later. For the probability of one unit along the

time sequence being identified with its counterpart depends on the type and number of elements that have already appeared.

Since the sequences of phonemes appear, not in space but in time, the position of a phoneme in one of these temporal sequences can be weighted. If, instead of reading and translating the texts, one were to listen to them, as do interpreters, the similarity of elements at the beginning of words or word groups becomes more important that it does at the end. This importance is also reflected in the natural tendency to form oral abbreviations like math and gym - rather than matics and nastics. The operative factor, however, is not sound, but sequential position. Some computerized systems are able to operate - even with proper names - with no more than five letters per word - provided they start counting from the first letter. It is almost as if a series of pictures were to appear in sequence, giving us more and more information as the film unrolls. In this sense, language is essentially kinetic, and if all its characteristics are to be measured, this must be taken into account.

2.3 Distance as Difference and Conversion: D (A~X), D (A→X)

We may regard the distance between languages either as the sum of their differences or as the amount of change needed to convert one language to the other.

In adding the differences between two languages, one may conceive them as opposing forces, or vectors, which can be combined as a measure of divergence (differential or vectorial distance). Or one can regard these differences as the amount of work or number of operations necessary to convert one language into the other (conversion distance). The first, being a combination of forces, is non-directional; the second, being a transfer of operations, is uni-directional. Let us examine what the measurement of each involves.

2.3.1 Differential Distance: D (A~X)

The measurement of the differential distance between Language A and Language X, being non-directional, is obtained by the summation of differences between both languages, so that D(A~X)= A - X + X - A. For example, in measuring the difference in the system of determiners in French and its counterpart in English, the fact that the single indefinite article (a) is equivalent to the two of French (un, une)is the same amount of difference as the fact that the person determiner in English (his) is equivalent to two forms in French (son, sa), a difference of one in each case. To say that the English (his,her)= French (son, sa) and count a distance of zero would be to ignore the fact that (her = son, sa) - (1-2) and (his = son, sa) - (1-2), and (son - his, her) - (2-1), (sa = his, her) - (2-1). So that the distance is not zero, but rather, in natural numbers: (1-2) + (1-2) + (2-1) + (2-1) = 4, that is a vectorial distance of 4. The vectorial distance between the series of singular determiners in English and their French equivalent will then be a total of 11, calculated as follows:

E.	a	the	this	+	that	my	your	his	+	her
F.	un+une	le+la	ce+cette	ce+cette	ce+cette	mon+ma	ton+ta+votre	son+sa	son+sa	son+sa
	1	1	1	2	1	1	1	1	2	1
	2	2	2	1	2	2	3	2	1	2
	1 +	1 +	1 +	1 +	1 +	1 +	2 +	1 +	1 +	1 = 11

In other words, if we add the differences, between English and French, in the distinctions recognized in this series of determiners with those between French and English, we get a total of 11. If we were to add all such differences in all the semantic and formal elements of these two languages we would get an idea of

how far apart they are. But we would not know how many more distincitions would have to be made by a Frenchman learning English in comparison with those which an Englishman would have to make if he were learning French. Here the starting point would make a difference, since the distance between A and X would not be equivalent to the distance between X and A. We would have to know which language features are being converted into which.

2.3.2 Conversion Distance: D (A → X)

It may make a practical difference whether a learner of a second language starts off with a native language that makes a large number of compulsory distinctions and tries to master one that does not, or whether he is going from a simple language to a more complex one. In order to find out this practical difference, differential measures would not be sufficient. What is needed is an idea of how many extra distinctions have to be made when passing or converting from one specific language (Language A) to another (Language X), in other words, we would have to measure the conversion distance.

The measurement of conversion distance, being uni-directional, is obtained by computing the number of distinctions not present in the other language. In the above example of the set of singular determiners in English and French, it will be seen that for every English determiner there are at least two French ones, one masculine and the other feminine; so that when converting from English into French, there is additional distinction of gender each time. In addition to this, there is, in one determiner (the second person singular) an extra social distinction to be made between the general and the familiar forms. The conversion distance is the sum of these distinctions that do not exist in the source language; difference between the language to be attained (the target

language)(T) and the language already acquired (the departure or source language) (S). So that, S → T = T - S = A - X. Compare:

A(Target):	F.	un+une		le+la		ce+cette		ce+cette		mon+ma		ton+ta+votre		son+sa		son+sa
		2		2		2		2		2		3		2		2
	-	1		1		1		1		1		1		1		1
X(Source):	E.	a		the		this		that		my		your		his		her
	=	1	+	1	+	1	+	1	+	1	+	2	+	1	+	1 =9

D (sg.det) (E → F) = 9

These represent distinctions of only two types, one (masculine-feminine) with the highest productivity, covering all or almost all determiners and nouns, and the other (ton-ta, votre) limited to the second person singular determiner and pronoun.

Measurement in the other direction, however, gives quite a different picture. The French speaker learning English has to make only a few extra distinctions he does not make in his French determiners, but these are in different semantic areas. So that, F. → E. = E. - F. :

A(Target)	E.	a		the		this + that			my		your		his + her			
X(Source)	F.	un une		le la		ce cette			mon ma		ton ta		son sa			
		1		1		2			1		1		2			
	-	1		1		1			1		1		1			
E. - F.	=	0	+	0	+	1	+		0	+	0	+	1	=	2	

D (sg.det) F.→ E. = 2

A French learner of English would have to make two extra distinctions which he does not make in French, and each of a different type, the first having to do with proximity and the second with sex. Both are high intensity items, however, because of their frequency of occurrence but since they apply to small classes of

words, their item productivity in the code is lower (see below).

2.4 Taxonomic Distance and Integral Distance: $\dot{D}$, $\bar{D}$

One can measure the distance between languages either by taking the differences category by category, that is, taxonomically, or by attempting to abstract from them integral measures in order to obtain an over-all picture.

In the taxonomic measures we have to deal with classes, categories or dimensions. Each has to be considered separately and measured by a procedure appropriate to the dimension being analysed. For example, differences in syntactic relationship such as the concord between adjective and noun, may not be measured in the same way as the difference in class membership where the item may be an adjective in one language and a noun in the other.

In the integral measurement of interlingual distance, only the equivalents as a whole, and only as they meet the eye or the ear, are measured for their observable differences.

In the first case, the results of the measurement appear as a sequence of figures each representing the distance in the category or dimension included. In the second, the distance appears as a single integrated figure.

2.5 Semantic Distance and Formal Distance: $D^{(A-X)}$, $D_{(A-X)}$

Studies of language have kept separate the description of what is said from that of how it is said. The dichotomy has variously been called semantic vs. formal and content vs. expression. Without denying that all language features, even those of meaning, are formalized in the code, let us follow the conventional usage by calling what is said, the content form, 'semantic', and how it is said, the expression form, 'formal', for want of a better term.

It will be seen that formal differences between languages are not equivalent to semantic differences. A look at a few deceptive cognates will make this

clear. For example, the form location is the same in written French as it is in written English. In the semantic code of French, however, it does not cover the same area of meaning as it does in English, since it would sometimes have to be rendered in that language as rentals, to let, and other forms. An examination of the semantic equivalents in English and French of such words as administration, information, relation, transmission, page, machine - to name only these - would reveal the same distinction between formal identity and semantic divergence.

The distinction, therefore, may be an important one, since the grouping of concept categories is not necessarily connected with the relation between word forms. The fact, for example, that languages group words into masculine and feminine does not signify that the difference has anything to do with the meanings of these words.

In comparing language codes, therefore, the distinction between semantic distance $D^{(A-X)}$ and formal distance $D_{(A-X)}$ will have to be maintained. In measuring the distance in discourse, however, the distincition no longer holds, since the comparable texts being equivalent, the semantic component is a constant.

2.6 Direct Distance and Indirect Distance

Interlingual distance may be direct or indirect. If the difference between two languages or dialects can be measured directly, there may seem to be no point in proceeding otherwise.

In certain studies, however, where interlingual measurement may be necessary, the only significant differences may be indirect ones. For example, if we wish to measure the difference in standard of the two languages in a bilingual community, we will not find the answers by comparing the one with the other, but rather by measuring each against its standard regional or national forms.

Or in the study of two mixed languages with a common parent, we can find out by indirect measurement the extent of divergence of each and the results of the different types of mixture.

Measurement of interlingual distance may therefore be done either directly (D = A - X) by comparing one language with the other, or indirectly by using an intermediary (J) language or dialect, as is sometimes necessary in practice: D = (A-J) - (X-J).

2.7 Diversity, Intensity and Productivity

In attempting to measure the distance between languages, we may wish to know, not only how different the languages are, but also how important are the differences, both to their users and to the workings of the languages themselves.

The distance between any two comparable sets of features in two languages may vary in three respects - in diversity, in intensity and in productivity. Diversity has to do with the number of distinctions a language makes in its grammar, its vocabulary and its phonetics. Intensity is a measure of the importance of such distinctions. Productivity shows the number of areas (situations, structures and categories) to which the distinctions apply. Distance between languages can be measured as the sum of the differences in diversity, intensity and productivity of their comparable elements.

Differences in diversity can be measured by subtracting the number of constituents of one language from their counterparts in the other. For example, in the semantic area of the third person singual pronouns, the difference between English and French is 1: (he, she, it) - (il, elle) = 3 - 2 = 1.

Differences in intensity can be measured by the use of the appropriate scale of importance , depending on whether the comparable units are discourse features like the interconnectives (although, since, if, ...) logical abstracts (be, seem....) or environmental variables (wine, snow). Appropriate scales

may be found in the indices of frequency, range, coverage, availability and familiarity.

Productivity has to do with the degree of generality at which the distinction is actually applied. Does it affect a large class of elements (like all sentences, or all nouns), or only a small class? Within the class, does it affect all members or only a few (as do the foot-feet, goose-geese plurals in English)? French, for example, makes action on self vs. action outside self a compulsory distinction, whereas English permits a choice (F. Je me lave./ E. I wash, I wash myself). This potential of difference in compulsory distinctions can be multiplied by the number of items so affected. Differences in productivity can also be regarded as the number of structures or units into which an element can fit. For example, in the formation of plurals by vowel mutation (e.g. E. mouse-mice/ G. Maus-Mäuse), German is more productive than is English, where the plurals are usually in - s. In this respect, English is closer to Spanish than it is to its Germanic cousin. Productivity differences, therefore, can be measured by counting the number of items in each language affected by the difference, and by calculating the number of possible utterences which the distinctions may produce.

In sum, in measuring the extent to which different languages differ, we must distinguish between what we are measuring and where we are measuring it. What we are measuring may be the differences themselves, or the amount of work needed to change one language into another (differential and conversion distances). Where we do the measuring may involve us either in comparing code items or in analysing samples of speech (code and discourse distances). The techniques of measurement used will depend on the types of distances measured.

3

DISTANCE BETWEEN CODES

3.1 Semantic Differences

3.1.1 Universals of Meaning

3.1.2 Semantic Measurement

3.1.2.1 Grammatical Meaning
3.1.2.2 Lexical Meaning

3.2 Formal Differences

3.2.1 Universals of Form

3.2.2 Formal Measurement

3.2.2.1 Grammatical Form
3.2.2.2 Lexical Form
3.2.2.3 Phonetic Form
a) Prosodic Distance
b) Allophonic Distance

3.3 Limitations of Intercodal Measurement

3. DISTANCE BETWEEN CODES

In measuring the differences between the codes of two languages it may be useful to separate the semantic from the formal, since each poses problems of a different nature.

3.1 Semantic Differences

Before attempting to measure the semantic differences between languages, it is important to consider the problems of analysis in the vocabulary and grammar of the languages in relation to possible semantic universals.

3.1.1 Universals of Meaning

The distincitions which a language makes are not limited to grammatical categories. The concept categories of languages are so varied and overlapping that it is difficult to divide them into grammatical meanings and lexical meanings. One can imagine the possibility of arranging them into scales of inclusiveness in each of the areas of logical environment, including the naming of things and relationships, actions and states of being that all languages must presumably have. Communication, a compromise between a desire to include and the need to exclude is also a consideration in the evolution of all codes. For all codes must make some sort of compromise between accuracy, efficiency and redundancy. Language codes have ways of taking the audience into account with register terms and intonation patterns for the transmission of semantic overtones. Each language must also adapt to changes in the environment with which it has to deal. It does this through a process of variation of its categories by analogy (hands of a clock) and combinations (G. Stechnadel) and collocations (as a matter of fact, on account of).

Through means that it may have at its disposal, each language must be able to label objects (parts of the body, food, clothing, shelter) properties of

objects (number, color, sound, shape, odor, taste), actions (eating, sleeping, walking), relationship (in, on, at, by, for), operations (identification of self and others, questioning, answering).

Also by the above processes of analogy and combination, each language in its own way - and often in seemingly arbitrary fashion - associates certain meanings (semantic features) to one or the other of these labels. So that the same group of features (e.g. thunder) may be associated in one language with the category of things, in another language with the category of actions, and in still another language with the category of qualities.

Keeping in mind the fact that discovery of features that all languages do (not must) have may well increase any measured distance between languages, can we still proceed to measuring relative distances in the area of logico-grammatical categories where such universals are most likely to be found? If one were to arrange the semantic distinctions into sets;i)there is likely to be a continual overlap of sets, and ii) what is more inclusive in one language (A) may be less inclusive in the other language (X). For example, French does not distinguish between persons and things in its pronoun system (il, elle) whereas English does (he +she, it). But in the area of questions in French (qui, quoi) and in English (who, what) both languages make the distinction. In relative connectives, however, only English makes it (E. The person who....The thing which...../ F. La personne qui....La chose qui.....)

When we consider the differences in the distinctions which languages make, the arbitrariness in which they group them, and the indeterminancy of the categories they create, how can we attempt to measure the semantic differences between languages?

3.1.2 Semantic Measurement: $D^{(A-X)}$

Although there have been attempts to study the measurement of meaning, semantic measurement, even for a single language, poses some formidable difficulties. To begin with, there is still no agreement on the meaning of meaning. To a psychologist, it may mean what one feels about a word; to a philosopher, it may mean what one thinks about it. Psychologists have attempted to measure differences in affective meaning by comparing the patterns of word association derived from testing different samples of a population; the responses have been arranged in such a way as to bring out what is known as the semantic differential. Logicians, on the other hand, have tried to reduce meaning to logical categories.

While recognizing the legitimacy and usefulness of such studies, one must, in comparing different codes, consider the linguistic rather than the affective or logical differences in meaning, and try to propose ways by which measurements might be elaborated. And although it is difficult to draw a line between lexical and grammatical meaning, it is necessary to begin by recognizing these conventional and convenient distinctions, since they suggest different approaches to the problem of measurement.

3.1.2.1 Grammatical Meaning

When a speaker puts semantic categories together to produce an utterance, he may have to take other semantic features into account - and these may vary from language to language. The concepts indicated by the word table, for example, may cover just about as much ground in English as it does in French. But when it is referred to a second time with a substitute word (it/elle), the English speaker must make a compulsory semantic distinction pointing it out as a thing; whereas the French speaker makes only a formal distinction,

repeating the feminine classifications of the article (la, une). It is as if a newcomer into a strange land had to learn not only that the land is inhabited by different categories of persons, recognized either(semantically) by what they do, or (formally) by how they look or dress. He must also learn whom to ask for what.

How can one measure this semantic compulsion to which varying languages submit their speakers? The problem is to identify the compulsory distinction that must be made in each language. Some of these are obvious, like the distinction between positive and negative in English; others are less so, as is the progressive in English and the subjunctive in French. It is important to keep in mind that we are here treating only the semantic features - grammatical meaning, as it were, and not grammatical form. The distinction is important, since apparently equivalent features may be semantic in one language and formal in the other - and the fact that they are, is likely to increase the distance between the languages. For example, both English and French have possessive determiners (E. his, her/ F. son, sa); but in English the masculine-feminine distinction is semantic; in French it is formal, agreeing,not with the owner but with the form-class (gender) of the word with which it is used, - son jardin/ his garden, her garden; sa maison/ his house, her house. In other words, English, unlike most Indo-European languages - with the possible exception of Modern Persian and Armenian - has a natural gender (semantic), while French - like most languages in the family - has a formal gender. But formal gender is likely to vary from language to language, making the sun masculine in French (le soleil) and feminine in German (die Sonne)and the moon feminine in French (la lune) and masculine in German (der Mond). And the classification is likely to seem unreasonable to the stranger. Why should a soldier in French (le soldat) become feminine when he acts as a sentinal (la sentinelle)? And why should young girls and old women neuter in German, (das Mädchen, das Weib)?

It is perhaps possible to divide the compulsory semantic restraints of grammar into categories related to grammatical function - such as identification - and to predication. The components of these categories, however, may remain incomplete, since theoretically any language can contain any obligatory type of semantic distinction. That is the significance of the slot labeled 'other' (see Table 2). As a category of predication, for example, 'other' could represent the distinction between verbs of activity and those of displacement - compulsory in French, but not in English. Compare:

E. At home, he walks. / F. Chez-lui, il marche.

E. He walks home. / F. Il va chez-lui à pied.

In the categories of identification and function, there are also other possibilities that may be difficult to predict *a priori*. The identification of positive, as distinct from negative, partitives is found in English but not in French (E. some-any/ F. de).

Once a complete inventory of these distinctions is obtained (and we do not yet have such an inventory) we can use the presence or absence of such compulsory forces as vectors in measuring the difference between two languages (See Table 2).

TABLE 2
UNIVERSAL CATEGORIES OF LANGUAGE CODES

I. CATEGORIES OF FUNCTION

	statement question	actor action	action goal	qualifier qualified	relator relation	other
A						
X						

II. CATEGORIES OF IDENTIFICATION

	expression reference	addressor addressee	one many	definite indefinite	human animal	masculine feminine	general honorific	internal external	container contained	animate inanimate	other
A											
X											

III. CATEGORIES OF PREDICATION

	positive negative	past non-past	future non-future	perfect imperfect	certain possible	simple conditional	motion non-motion	active passive	freedom obligation	other
A										
X										

There are some disadvantages in the use of a general framework of possible grammatical distinctions. First, there may be difficulty in making sure that the framework contains all possible distinctions. Secondly, some languages make a number of distinctions at the same time; and one is never sure which and how many are intended. There are, for example, about forty possible semantic distinctions of case, that is, some forty ways in which one thing (or nominal concept) may be related to another. In Latin, the nominative case covers one, the accusative, two and the dative four.

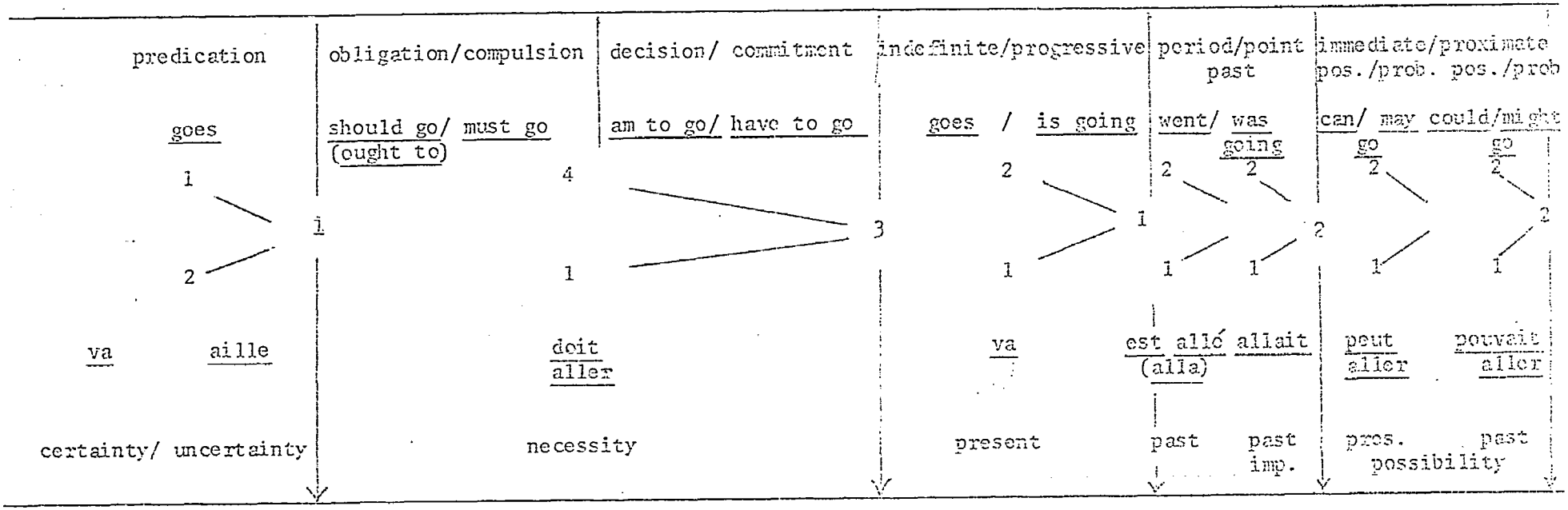
Instead of proceeding from a check-list of all possible semantic distinctions in the grammar of languages, it is possible to adopt the less ambitious procedure of comparing the two grammatical systems from the semantic point of view. In doing so, however, one must be careful not to confuse form with function on the one hand, nor constraints with alternatives on the other. A single form as we have seen may express several relations of different types - person, number, case, tense, mood, voice. For example, the - o in Lat. amo (I love) expresses the present indicative of the first person singular of amare; the single form -atur in amatur (He, she or it is loved) expresses distinction of voice, person and tense.

A constraint cannot be treated simply as an alternative. Take the forms of the French verb as shown in the words a fini, finit and finissait. There may sometimes be three ways of viewing past time; but only two may constitute a semantic constraint, since the speaker of French may ignore the distinction expressed by the simple past and get along quite well with the perfect and imperfect tenses.

Keeping these precautions in mind, therefore, we can isolate from the two languages the cases where more semantic distinctions have to be made in the

grammatical forms of one language than in the corresponding forms of the other. This is equivalent to the measurement of the conversion distances in the grammars of both languages - measuring in both directions, the difference being $D^{(A \sim X)} = (A \rightarrow X) + (X \rightarrow A)$. From this point of view, then, let us take a look at the verbal systems in English and French. (See Table 3.)

TABLE 3
SEMANTIC DIFFERENCES IN ENGLISH AND FRENCH VERBAL SYSTEMS



Not all these differences, however, can have the same value within the respective codes. Regardless of frequency of usage, the degree of compulsion is not the same. One could argue that one of the differences in the codes between English and French is that the former has maintained the Indo-European dual in its category of number; but the forms in which it is found, although frequent, are residual. All languages must have such residual forms, since all living languages evolve. These forms will come out in the measurement and will have to be weighted for their productivity. It is true that the Indo-European number category did have a dual - for example, in Sanskrit (Sk. táu) - as well as a singular (Sk. sá) and a plural (Sk. tế). Although this dual has disappeared in French, it remains as a residual form in English. Compare:

You can't take both,	On ne peut pas prendre les deux,
but you can pick the better.	mais on peut choisir le meilleur.
You can't take all,	On ne peut pas tout prendre,
but you can pick the best.	mais on peut choisir le meilleur.

In the last analysis, grammar is simply a means for speakers to establish relations between meaningful elements of their discourse. What is coded as grammar, and what is not, cannot always be an either/or proposition. A grammar of English, for example, will vary according to how one classifies the words get and keep. If they are considered like have and can, as having grammatical functions, then we would have to add two more aspects to the system of the English verb, since these function words can combine with perfective (-ed) and imperfective (-ing) grammatical forms to produce what could be called grammatical meaning, which would have to be rendered in French by differences in mood, voice, or vocabulary.

Compare:

	Inceptive		Durative	
	Perfect	Imperfect	Perfect	Imperfect
E.	get married	get going	keep married	keep going
F.	se marient	se mettent en route	restent mariés	continuent

Since many semantic grammatical categories exist mostly in the minds of the categorizing grammarians, it seems advisable in measuring the semantic distance between languages not to rely entirely upon them. It is necessary to see how many of the differences can also be covered by concept categories considered as vocabulary.

3.1.2.2 Lexical Meaning

Since measurement is not description, available inventories of semantic features may be useful in establishing the indices of difference. Many of these inventories - even the most abstract and grammatical - can be found in certain unilingual and bilingual dictionaries - imperfect as they may be - supplemented by the descriptive and differential grammars of the two languages.

Because the groupings of distinctions of one language may overlap a number of groupings (categories) of the other language, and vice versa, it is necessary to measure the differences from both viewpoints: $(A \sim X) = (A \rightarrow X) + (X \rightarrow A)$. And since even the most complete dictionaries and grammars differ on what is category of meaning and what is usage, in referring to this valuable source for the measurement of semantic distance, it will be necessary to include both.

For example, if one takes the concept of man - which may perhaps be universal - and see what two languages, like English and French, group into it, one soon notices that although the groupings have a lot in common, there are remarkable divergences. The concept (man) does not cover the same number of areas of meaning, and in each area it does not cover the same specific uses. For each concept, the degree of congruence can be used as a measure of closeness or distance.

Since an inventory of this concept and its uses may be found in a modern bilingual dictionary, we shall use the same dictionary (e.g. Harrap's English-French/French-English Dictionary, Rev. ed. 1967) to measure the differences. Following the categories of the dictionary, we find that it recognized six categories for man (including one subcategory) (A=6) and four for homme, (X = 4), making a total of ten. Only two of these categories (mankind, and adult male) are identical, (A ≡ X) = 2., giving an overlap of 20%.

Not each category, however, has the same number of usages or 'meanings', and in these there is also a possibility of overlapping. It is therefore wise to count in both directions, i.e., the cases in which man = homme and homme = man. For example, although the category adult male (man/homme) exists in both languages, this does not mean that usages within the categories are always equivalent. In some cases they are (I'm your man/ Je suis votre homme) and (Soyez un homme/ Be a man); in other cases they are not (Good man!/Bravo!) and (Homme à femme/ Lady killer). If we add together these cases in which languages are not equivalent, we get another measure of difference.

In the six categories of man (translating each usage of English into its French equivalent) we have 15 cases (a = x) of equivalence and 42 cases (a ≠ x) of non-equivalence in a total of 57 usages listed. Contrariwise, in the four

categories of homme, this time translating from French into English (F/E), we find 12 cases (x = a) of equivalence as against seven (a ≠ x) of non-equivalence (F/E) out of a total of 19. If we add all these differences and express them as a percentage of the total number of categories and constituents, we get a semantic distance of .62, by applying the following weighted formula:

$$\dot{D}^{(A-X)} = \frac{1}{2}\left(\frac{\sum \text{Cat.}(A \neq X)}{\sum \text{Cat.}(A+X)} + \frac{\sum (a \neq x) = (x \neq a)}{\sum (a-x) + (x-a)}\right)$$

$$\dot{D}^{(\underline{\text{homme-man}})} = \frac{1}{2}\left(\frac{6}{9} + \frac{49}{76}\right) = \frac{1}{2}\left(.6 + .64\right) = \frac{1.24}{2} = .62$$

Since dictionaries and grammars, however, are loaded with low-frequency, unproductive, rare and obsolescent forms, it seems necessary to select those concepts which represent that part of the code actually in use, or the more general, universal or usual concepts. By way of example, we have selected ten general concepts and have arrived at some sample measurements of the distance between them in English and French (See Table 4.).

Identity of meaning, however, does not indicate identity of form no more than the degree of semantic difference indicates a corresponding degree of formal difference.

TABLE 4

SAMPLE SEMANTIC DISTANCES BETWEEN ENGLISH & FRENCH LEXICAL CODES

man / homme	62.0%
time / temps	73.4%
way / façon	89.9%
state / état	73.3%
world / monde	72.7%
life / vie	65.9%
come / venir	75.4%
see / voir	65.4%
high / haut	71.6%
house / maison	74.6%

Semantic distance of ten lexical items in English and French calculated by percentage of two-way differences in concept coverage of an English word and its French dictionary equivalent.

3.2 Formal Differences

The formal features of language are more evident than the meanings they convey. The basic needs of communication and the limits of the human speech organs combine to place the formal features of languages within certain boundaries. Let us discuss these universal formal features before elaborating ways of measuring their differences.

3.2.1 Universals of Form

The needs of communication require a limit to the length of an utterance necessary to convey a conventional semantic feature - and the more frequent or usual the feature, the shorter it is likely to be. But there is also a limit to the shortness, since another basic need of communication is that the messages reach their destination (the listener's mind) intact despite the distortions of human and therefore imperfect production and the transmission losses due to noise and distance, so that if one feature is lost another can take its place. In other words, there must be a number of redundant features in any language system. All systems of all natural languages seem designed to produce redundancy. And this redundancy is often constituted in the "useless" constraints which languages oblige their speakers to use. They appear in the grammar of languages as rules which seem pointless and arbitrary to the learners, rules like the use of both preposition and case ending to convey the same idea - (G. in seinem Beruf - trotz seines Berufs; L. a mari - ad mare) - but which must be obeyed.

In order to be able to create utterances, all languages must have some sort of elements - such as words and speech sounds, which can be said to form part of the code of the language. All languages, in order to become means of communication, must make some sort of distinction between these elements in the

code and the utterances which they can produce. Otherwise a language would be simply a collection of ready-made utterances, quite incapable of dealing with the complex needs of man. The potential number of utterances must therefore be greater than the number of elements in the code of the language. In order for this to be possible, some elements must combine in different ways to produce different utterances. For example, the utterance 'dog bites man' conveys a different message than does the utterance 'man bites dog'. The different ways in which the elements combine are the structures of the language, each structure containing types of elements in certain positions, forming the units of the structure.

These units are the immediate divisions of the structure into its components. In some languages, clauses are the immediate divisions or units of the sentence structure, phrases the units of clause structure, words the units of phrase structure, etc. In this way, some structures are more inclusive than others, and this inclusiveness forms a hierarchy of inclusion which all languages contain, since the minimal operational number seems to be three, e.g., sentence, word, sound.

The necessary difference between code and message supposes that a structure must be able to produce a number of different utterances by using different combinations of units (dog bites man, dog bites bear, bear bites man, bear kills man, etc.). The more different utterances produced, the greater the productivity of the structure. This depends on the number of components of the unit which can occupy the same position, that is, it depends on the depth of the 'well' into which the structure can dip, as it were, to produce new combinations by filling its blanks. If the position is occupied by a determiner, (this, that, the, his, her, etc.), for example, the depth will be less than if occupied by an adjective

(good, bad, tall, blue, etc.). The depth of a class of units in one language is not always the same as it is in another. Finally, length, class and relationship of the units obviously vary from language to language.

To sum up the universal formal characteristics of languages, therefore, we can state the following:

i) All languages contain elements.

ii) These elements comprise units and structures.

iii) Some of these are more inclusive than other, making all languages essentially hierarchical.

iv) The units appear in the structures in a certain order.

v) Each unit has some length.

vi) Each class of unit has one or more members, giving the class depth.

vii) Each unit is in some way connected with other units, giving it a set of possible relationships.

The formal elements of a language are not necessarily determined by the way the language categorizes reality. The fact, for example, that two languages distinguish persons from things does not mean that they use the same sort of formal elements to do so. One may do it through its determiners; the other through formal non-classes or word endings. The semantic potential of a language - its capacity to express and combine categories in various levels of meaning (logical, grammatical, environmental, interpersonal and intratextual) is one thing; its formal potential is quite another.

The formal potential of a language is revealed in the means or choices it has to generate the sort of meaningful messages of which its semantic system is capable. These choices operate both vertically and horizontally, as it were, selecting the appropriate structure down the hierarchy while bringing in the appropriate units to combine into a linear sequence which is transformed by the speaker into an utterance. It may be that the higher the level in the

hierarchy, the less the difference between languages; the most abstract choices may be universal. But at the bottom of the ladder, at the level of the sequence of sounds constituting the utterance, the differences between languages can easily be seen and heard. They constitute the other end of the continuum where all languages must again find an outlet in the physical world.

3.2.2 Formal Measurement

The formal differences between language codes can be measured through elements that are comparable. A word or form in one language can be compared with a word or form in another language, a sound in one language can be distinguished from a sound in another language. It is therefore possible to separate grammatical form from lexical form and phonetic form.

3.2.2.1 Grammatical Form

Although the vocabulary of two languages may be highly similar in form and meanings, the way it is put into operation may require quite different activities in one language than it does in the other. Although the form police looks the same and means much the same thing in one of its usages in French as it does in English, once the English speaker decides to use this word in a French sentence, he has to take into account the grammatical differences between the two languages, not only grammatical meaning, classifying the notion as singular rather than plural, but formal aa well, making them (or it) feminine instead of masculine (la police); and he has to maintain this formal distinction whenever he adds an adjective (la police italienne). It is as if our newcomer in a strange land sees familiar objects related to many unfamiliar persons, and has to decide between the persons; so that he is forced to make a choice between types and relationships he did not know existed. He must at the same time learn what they are, what they contain and what to do about them.

	le			chez-elle
police?		italien	va	?
	la	italienne		chez-lui
			vont	

If his language is not too different, he may be familiar with the type of distinction, without knowing the distinction itself. For example, if the stranger to French were Italian rather than English, the word police might not look quite so familiar to him, since the Italian version is a bit different (polizia); but he will be familiar with the process of formal word gender. What he will have to learn is the specific gender of the French equivalents. And even some of the close cognates may mislead him. In this case, he is likely to be right three out of four times and wrong one out of four times, since the difference in the gender of Italian-French cognates has been estimated as about 25%. The English speaker, on the other hand, has only a fifty-fifty change of getting the gender of his cognates of French. But he first has to get used both to the idea of formal word gender and also learn which word is which gender. This leads to the idea of level or hierarchy in distinguishing formal grammatical categories. And it suggests that it is unwise to add a distinction at one level with a distinction at another until levels have been normalized in scale.

If we start with the most general formal distinctions possible, like the actor-action distinction, and by a continual sub-classification process, work down to the degree of differentiation in word-forms, we can keep the levels separate and then average the differences.

Our inquiry would go something like this: Are there formal distinctions between actor and action and goal? What are they? In the

actor part of the utterance, what formal distinctions exist? In each of these (e.g. nouns) what formal distinctions are made (e.g. gender)? How are these distinctions made (e.g. affixation, relation words, word-order)? What relations are formally marked (e.g. determiner-noun)? How are they marked (e.g. agreement in form)? What is the extent of the relationship (e.g. 24 forms of determinatives in French as against 12 in English)? What is the degree of formal relationship (e.g. the common th- element in this-that and these-those)? Let us see how answers to such questions would produce quantifiable results. (See Table 5.)

These differences in the forms of function-word groups, after having been thus measured intralinguistically, are then measured interlinguistically, that is, against their corresponding opposites in the other language - the against le, la, and les, their against leur, etc. These differences, however, can be measured at the lexical level, along with the vocabulary of the languages, showing distance in lexical form.

TABLE 5

SAMPLE CHECK LIST OF FORMAL DIFFERENCES IN NOMINALS

	ACTOR-ACTION	ACTION-GOAL	OTHER	
A	✓	✓		
X	✓	✓		
A ~ X	0	0		0

	WORD ORDER	RELATION WORDS	INFLECTIONS	OTHER	
A	✓		✓		
X	✓	✓			
A ~ X	0	1	1		2

	NOUNS	DETERMINERS	ADJECTIVES	PRONOUNS	OTHER	
A	✓	✓	✓	✓		
X	✓	✓	✓	✓		
A ~ X	0	0	0	0		0

	NUMBER	PERSON	GENDER	CASE	OTHER	
A	✓			✓		
X	✓		✓			
A ~ X	0		1	1		2

	Markers	Affixation	Suppletion	Reduplication	Internal Change	
A		✓	✓		✓	
X	✓	✓				
A ~ X	1	0	1		1	3

Agreement:

	Determiner	Adjective	Other	
A	✓			
X	✓	✓		
A ~ X	0	1		1

	Number	Person	Gender	Case	Other	
A		✓		✓		
X	✓	✓	✓	✓		
A ~ X	1	0	1	0		2

Extent:

	the,	this-that,	these-those,	my,	your,	his-her,	our,	their	
A									
X	le la les	ce cette	ces	mon ma mes	ton ta tes	votre vos	son sa ses	notre nos	leur leurs
A ~ X	2	0	1	2	4	1	1	1	12

TABLE 5 (continued)

Concentration:

the	the	the	this	this	that	that	these	those	my	my	my	your	your	your	his	his	his	her	her	her	our	our	your	your	their	their
-	-	-	s	s	s	s	p	p	1s	1s	1s	2s	2s	2s	3s	3s	3s	3s	3s	3s	1p	1p	2p	2p	3p	3p
sm	sf	p	sm	sf	sm	sf	p	p	1sm	1sf	1sp	2sm	2sf	2sp	3sm	3sf	3sp	3sm	3sf	3sp	1ps	1pp	2ps	2pp	3ps	3pp
le	la	les	ce	cette	ce	cette	ces	ces	mon	ma	mes	ton	ta	tes	son	sa	ses	son	sa	ses	notre	nos	votre	vos	leur	leurs
0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	1	2	2	2	2	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2 +	2 +	1 +	1 +	1 +	1 +	1 +	1 +	0 +	0 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 +	1 =27

s = singular, p = plural, m = masculine, f = feminine, 1 = 1st person, 2 = 2nd person, 3 = 3rd person.

Differentiation:

this these			that those								
			6/9		4/9						
le les	ce ces	cette ces	ce cette	mon mes	ton tes	son ses	notre nos	votre vos	leur leurs		
4/5	4/5	4/3	4/7	2/6	2/6	2/6	4/8	4/8	8/9		
ma mes	ta tes	sa ses	mon ma	ton ta	son sa						
2/5	2/5	2/5	2/5	2/5	2/5						

3.2.2.2 Lexical Form

The most obvious of all language elements is the word. An unsophisticated speaker of a language can recognize some of its words, although he may not be able to isolate the syntactic, morphological or phonetic elements. An uneducated language user can also recognize similarities between word-forms of different languages (e.g. G. Mann and E. man). He does so not only by the type and number of similar sounds or letters (e.g. time will not be identified with emit, although both words have the same type and number of letters). He must also recognize a similarity in the sequence, pattern or configuration of the units, their order and position (e.g. time and tempo will appear as more similar than time and emit). These similarities in configuration can be seen or heard by making a linear juxtaposition of one component in one language with its counterpart in the other. The question is one of sequence rather than of components. There are two types of linear correspondence - positional and equivalent. Positional correspondence is the one-to-one juxtaposition of elements in two interfacing units according to their positions.

For example: E. c a t
F. c h a t.

Equivalent correspondence involves the interfacing of two units according to the similarity or identity of their constituent elements. There is maximal equivalence when the interfacing elements are so juxtaposed as to achieve the optimal number of identities without changing their sequential order. For example:

E. c a t
F. ch a t

This latter arrangement is the appropriate measure for comparing the content of units; the former is suited for the measurement of their structural difference (See below). Using one type of interfacing instead of the other is likely to make a difference in the results of the measurements. For example,

English cotton and Spanish algodón, being cognates, have a certain degree of formal similarity. Which of the two measures is likely to reveal this? Compare:

Positional Correspondence	Equivalent Correspondence
E. c o t t o n	E. c o t t o n
S. a l g o d ó n	S. a l g o d ó n
All pairs different.	Three pairs identical

In other words, one calculation makes the words entirely different in form; the other gives their difference as being a little more than half.

In the cases of two languages with highly unphonological spelling systems, such maximal visual correspondence may make some word-forms appear more similar than the unsophisticated speaker of either language would estimate. For example:

F. h o m m e
E. m a n

This is avoided if the interfacing is done on the basis of constituent syllable units.

E. c o t t o n	E. m a n
S. á l g o d ón	F. h o m m e

Syllabically the English monosyllable time can be interfaced with the Spanish di-syllable tiempo:

t i m e
t ie m p o

One would not then identify the -e in time with the -e in tiempo. Once two forms have been thus juxtaposed, or interfaced, it is possible to measure the formal distance between them. Distance in the written forms and the spoken forms can both be measured. Since the spoken forms, however, suppose the juxtaposition of equivalent sounds, which may differ in degree (e.g. the /t-/ in time and in tiempo are not pronounced in identical fashion) we begin with differences that can be seen, and leave until later the more complex measurement of differences that can be heard.

By a linear juxtaposition of two graphic forms, therefore, we can see whether or not two interfacing elements (letters) are the same or different. If they are identical, the distance between the two elements is obviously nil and can consequently be counted as zero. If they are different, there are two possibilities - either there are no interfacing letters, in which case the distance can be counted as one; or the interfacing letters are different, in which case the distance is counted as two. This can be reduced to a simple formula:

i) Interface for maximal syllabic equivalence.

ii) Count combined total units (T).

iii) Check and count identical interfacing units held in common (C).

iv) Subtract these (C) from the total (T), giving a measure of the distance (D).

v) Express in relation to the combined length (T).

So that,

$$D_{(A-X)} = \frac{T - C}{T}$$

Going back now to our last example (E. time - S. tiempo), the distance would be 4, calculated thus:

t i m e — T (total units) = 10

or

t ie m p o — C (common units) = 6

0 + 1+0+ 2 +1 = 4 — Difference = 4

$$D_{(\text{time-tiempo})} = \frac{T - C}{T} = \frac{10 - 6}{10} = \frac{4}{10}$$

These differences may also be reduced to percentages or made to range between zero and one, a procedure that would be needed in adding or comparing differences. For example, we can see that the difference of 4 between time and tiempo, being 4 out of a combined total of 10 is .4 or 40%. For the next concept (man), the formal distance between the English form (man) and the French form (homme) interfaced syllabically, is 8 out of 8, that is 1, or 100%.

This reduction permits us to combine formal distance with semantic distance, likewise expressed in percentages, by showing the one below the other. In the above example, the semantic distance $D^{(A-X)}$ for (homme-man) was .62. So that,

$$D \begin{matrix} (\text{homme} \sim \text{man}) \\ (\text{homme} \sim \text{man}) \end{matrix} = \begin{matrix} .62 \\ 1 \end{matrix}$$

An even more accurate measure can be obtained by weighting each interfacing similarity according to its position in the linear sequence, since it may be argued that it is linearly that languages are perceived. But since this involves more complex calculations, we shall leave it until later (5.2.2 below).

Let us now extend our measurement by considering the difference between how the words have to be pronounced, since differences in lexical form can be measured either as to how the words look or as to how they sound. The latter will depend on our knowledge of code difference in phonetic form.

3.2.2.3 Phonetic Form

In measuring the distance between two codes, we have made a distinction between formal and semantic measures, the latter having possible reference to the physical world and expressing characteristics which are not limited to language. In pushing our formal measures further and further down the lingaistic hierarchy, we also end up in the physical world. For sounds, like meanings, are not limited to language.

The ways in which meanings or concept categories are designated in sequences of language sounds are limited by a number of the human vocal tract. Because these sounds are produced on a time sequence, there must be limits between one sound and another. Because of the way the human breathing apparatus is constructed, there must also be units related to breath. But how far can we postulate the generality of other phonetic features? We might be safe in postulating the universality of the syllable, the consonant and voiced sound, without going so far as to claim that all languages have stops, fricatives, labials or dentals.

It would perhaps be safer to speak of biological universals and say that all normal human beings have a mouth, a tongue and a

throat. And since all language sounds must be produced within the limited capacities of these human organs adapted to the production of sounds, their universal physical potential (mouth closure, tongue position, etc.) may be used as a means of measuring the difference between languages at the base level in the hierarchy of language units.

On these universal physical features of the human body, it should be possible to establish valid measures of difference in the production of speech sounds, since we can assume that the essentials of the vocal organs are the same for all men in all places, or that everyone at birth has the potential for learning any human language. By measuring the relative position of these organs in the production of each sound of a language, we can determine the extent to which it is different from any sound in any other language. What are these organs? They are those which change their position or shape to produce significant differences in speech - the tongue, the lips, the jaw, the muscles of respiration and the vocal cords. These organs together produce two types of physically measurable units - one acoustic and the other physiological. The first is the appropriate measure for the prosodic distance of syllables; the second, for the allophonic distance of speech sounds.

a) Prosodic Distance

If we listen to a language we do not understand, we are first struck by the rise and fall of the voice and the beats of sound varying in length and loudness. Some people learn to recognize languages they do not understand simply by listening to these features of rhythm and intonation.

One can hear the difference between the syllables of languages, since they vary in the level and direction of the tone of the voice and in the relative length and force of their constituent sounds. This is easily observed in the pronunciation of homophonic place names and loan words.

The name Canada sounds different in English than it does in French or Hungarian, not so much because the sounds are pronounced with a slight difference, as because of the force and length of the syllables. Compare:

English	C	a	n	a	d	a	-	loud/ weak/ weak
French	C	a	n	a	d	a	-	medium/ medium/ loud
Hungarian	C	a	n	a	d	a	-	loud/ medium/ medium

It is through such homonyms that we can begin to measure the syllabic (prosodic) distance between phonetic codes.

There are at least two ways that syllabic differences can be measured - instrumentally and taxonomically. A single instrument, such as the cathode ray oscillograph or the acoustic spectograph can reveal differences in syllabic length and tone, measurable within a difference of a single cycle of vibration per second. The disadvantage of such instrumental measurement is that it can use samples from only a limited number of persons, so that minute individual speech and voice differences tend to obscure the differences between the language codes. Instrumental measurement can, however, be a help in establishing usable scales and categories of relative difference.

If we examine the relative differences only, we can place them in a number of categories by using a series of such distinctive features as high and low, strong and weak, fixed and moving, thus establishing a number of distinct categories within each language for force, length, height and direction of the syllabic tone. For each of these four syllabic variables, we can establish a scale, giving a numerical value to each point (e.g. short = 0, medium = 1 and strong = 2). By using these values as vectors, we can subtract them one from the other in interfacing syllables, to measure their prosodic distance in natural numbers (See Table 6).

In measuring prosodic distance, therefore, we take four syllabic variants into account: force, length, direction and height of voice. (Catenation is accounted for in phonetic measurement.)

Force: A syllable in the stream of speech may appear louder or quieter than that of its neighbours. This is measured by giving a value of 0 to the weakest (so as not to make variations of force dependent on length, the adding zeros gives zero). So that it is entirely possible for a syllable to be longer and at the same time quieter than another. For the loudest syllable we attribute the value of (2).

Length: The individual length of the syllable is measured by giving the value of (1) to the shortest in the length repertory of the language and (3) to the longest, while giving the value of (2) to the others.

Direction: Comparing two syllables, it may be noticed that the tone can remain level throughout the length of the syllable, or that it changes in one, two or more directions during the period of time taken to utter the syllable. If the tone remains level, we give the value of (1) to the syllable; if it changes, we give values according to the direction and amount of changing.

Height: Listening to two different syllables, may reveal that the one is higher in tone or lower than the other. One can measure this by comparing the tones

syllable by syllable to see the extent to which they correspond. The value of (0) is given to the lowest tone and (5) to the highest (See Table 6).

For example:

Measurement of prosodic differences in the English and French prenunciation of Canada, gives the following results:

	CA		NA		DA		
RHYTHM							
Force							
E.	2		0		0		
F.	1		1		2		
Vectorial Distance	1	+	1	+	2	=	4
Length							
E.	2		0		0		
F.	1		1		1		
Vectorial Distance	1	+	1	+	1	=	3
INTONATION							
Direction							
E.	1		0		0		
F.	0		0		0		
Vectorial Distance	1	+	0	+	0	=	1
Height							
E.	3		1		1		
F.	2		2		1		
Vectorial Distance	1	+	1	+	0	=	2
Total Vectorial Distance							10

TABLE 6

PROSODIC VALUES AND NOTATION

(S = syllable)

RHYTHM (marked above the syllable)						
Force	weak	medium	strong			
symbol:	(unmarked)	$\overset{'}{S}$	$\overset{''}{S}$			
value:	0	1	2			
Length	short	medium	long			
symbol:	$\overset{\smile}{S}$	$\overset{-}{S}$	$\overset{=}{S}$			
value:	0	1	2			
INTONATION (marked before the syllable)						
Direction	static	falling	rising	rise-fall	fall-rise	3 changes
symbol:	$\|S$	$\backslash S$	$/S$	$\wedge S$	$\vee S$	$\wedge\vee S$
value:	0	1	2	3	4	5
Height	extra low	low	mid-low	mid-high	high	extra high
* symbol:	$\overset{\|}{S}$	$\overset{\|}{S}$	$\overset{\|}{S}$	$\overset{\|}{S}$	$\overset{\|}{S}$	$\overset{\|}{S}$
value:	0	1	2	3	4	5

*Direction symbols ($\overset{+}{|}\ \backslash\ /\ \wedge\ \vee\ \sim$) are written below the line $\underline{\quad}$(0), through the line $\underline{+}$(1), on the line $\underline{|}$(2), under the line $\overline{|}$(3), through top line $\overline{+}$(4), on top of top line $\overline{|}$(5).

b) Allophonic Distance

As one approaches the physical basis of human speech, our possibilities of accurate measurement increase. If the above type of vectorial measurement is applied to different speech sounds, these physical variables can be given scale values corresponding to differences that can be seen as well as heard. They can be seen directly or indirectly with the aid of such photographic techniques as palatography and cineradiology

Since any sound can be described by the action of the organs which produce it, the extent of such action can be given a value on a scale. With six scales corresponding to the action of the speech organs, all relevant speech sounds can be described (See Figure 1).

Each sound in each language can therefore be measured according to I (its point of articulation), II (the degree to which the organ affected is closed or constricted), III (timing of action of the two or more organs involved in production of the sound), IV (shaping of other organs involved in the production of the sound), V (the direction of the flow of air through the mouth or nose by the action of the velum) and VI (the action of the vocal cords). To each of these variables scale values can be given (See Table 7).

To measure the distance in the phonetic code of two languages, juxtapose all their phonemes and, using the scale, measure the degree of difference in their distinctive features. For example:

English /æ / is: I(fronted = 4), II(mid-open to low = 6.5),

4|6.5|3|0|0|1| III(continuous = 3), IV(non-constrictive = 0),

V(oral = 0), and VI(voiced = 1).

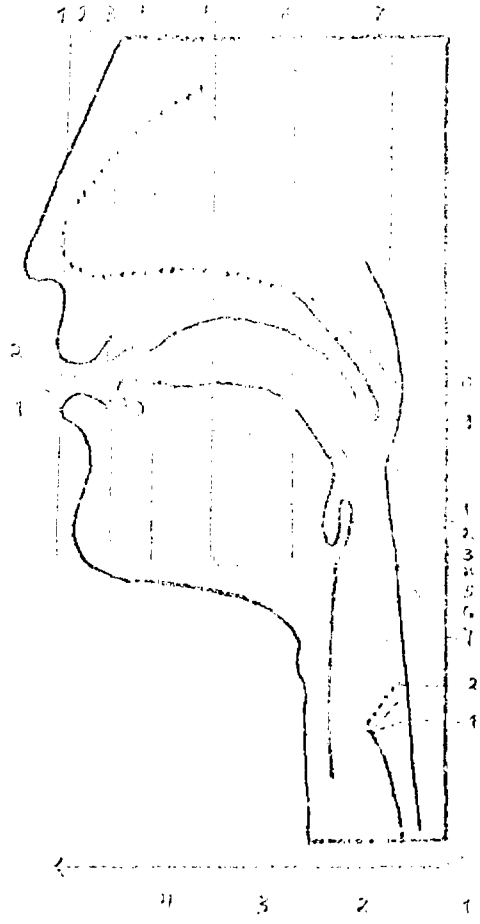
TABLE 7

VECTORIAL VALUES OF PHONEME COMPONENTS

Values	0	1	2	3	4	5	6	7
Scales I	-	bilabial	labiodental	interdental	apical/ front	central	back	glottal
II	-	closure	friction	semivowel	high	mid-closed	mid-open	low
III	-	no constriction	momentary	continuous	length-ened			
IV	no co-articulation	vocalic (off-glide)	affrication					
V	closed velum	open velum						
VI	-	vibrating vocal cords	vocal cords open					

Figure 1

SCHEMATIC VIEW OF PHONETIC VECTORS



French /a/ is:

4	7	3	0	0	1

I(fronted = 4), II(low = 7), III(continuous = 3)
IV(non-constrictive = 0), V(oral - 0), VI(voiced = 1).

Juxtaposing the comparable values and subtracting in natural numbers gives the vectorial distances between the two sounds, thus:

	I	II	III	IV	V	VI
E. / æ / :	4	6.5	3	0	0	1
F. / a / :	4	7	3	0	0	1

D /æ - a/ = 0 + .5 + 0 + 0 + 0 + 0 = .5

The vectorial distance between the two sounds is therefore .5 .

When all the sound features of one language have thus been juxtaposed with those of another, the result is a reference table to vectorial distances in the phonetic codes. By way of example, we have elaborated such a table to measure the vectorial distances between the sounds of English and those of French (See Table 8).

TABLE 8

ENGLISH-FRENCH PHONEME DISTANCES

8.1	English /æ - m/	-	French / a - o/
8.2	English /æ - m/	-	French / ɔ̃ - ɜ̃/
8.3	English /n - dʒ/	-	French /a - o/
8.4	English /n - dʒ/	-	French /ɔ̃ - ɜ̃/

TABLE 8.1 E.(æ - m)

ENGLISH / FRENCH

PHONEME DISTANCES

As vectorial differences in distinctive features

F.(a - o)

	æ	ɑ	ɔ	əi	əə	b	d	ð	ɛ	ʒ	ə	ei	f	g	h	i	k	j	ŋ	t	θ
	4	5.5	6	4	4	1	4.5	3	4	5	5	4	2	6	7	4	4.5	5	6	4.5	1
	6.5	7	7	7	7	1	1	2	5.5	5	6	5.5	2	1	2	4..	4.5	3	1.5	1	1
	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	3
	0	0	.5	1	1.5	0	0	0	0	0	0	1	0	0	1	.5	0	0	0	.5	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	2	2	1	2
æ	4 7 3 0 0 1	.5	1.5	2.5	1	1.5	9	6.5	6	1.5	3	2	2.5	8	8	10	8.5	3	5	8.5	9	10
ɑ	6 7 3 0 0 1	2.5	.5	.5	3	3.5	11	7.5	8	3.5	3	2	4.5	10	6	8	5.5	4	5	6.5	10	11
ɔ	5.5 7 3 0 1 1	3.	1	2	3.5	4	11.5	8	8.5	7	3.5	2.5	5	10.5	7.5	9.5	6	4.5	5.5	8	10.5	10.5
b	1 1 3 0 0 1	8.5	10.5	11.5	10	10.5	0	3.5	3	7.5	8	9	8.5	3	5	9	6.5	7	6	6.5	6	7
d	4 1 3 0 0 1	5.5	7.5	8.5	7	7.5	3	.5	2	4.5	5	6	5.5	4	3	6	3.5	4	3	3.5	5	4
e	4 5 3 0 0 1	1.5	3.5	4.5	3	3.5	7	4.5	4	.5	1	2	1.5	6	6	8	4.5	1	3	6.5	7	8
ɛ	4 6 3 0 0 1	.5	2.5	3.5	2	2.5	8	5.5	5	.5	2	1	1.5	7	7	9	8.5	2	4	7.5	8	9
ʒ	4 6 3 0 1 1	1.5	3.5	4.5	3	3.5	9	6.5	6	1.5	3	2.	2.5	8	8	10	8.5	3	5	8.5	9	8
ə	5 5.5 3 1 0 1	3	3	3	2.5	3	9.5	6	6.5	2	1.5	1.5	1	8.5	6.5	6.5	3	3.5	3.5	7	7.5	10.5
f	2 2 3 0 0 2	7.5	9.5	10.5	9	9.5	5	4.5	2	6.5	7	8	7.5	0	6	6	5.5	6	5	4.5	7	11
g	6 1 3 0 0 1	7.5	6.5	6.5	7	9.5	5	1.5	4	1.5	1.5	6	7.5	6	0	4	5.5	5	3	1.5	4	6
i	4 4 3 0 0 1	2.5	4.5	5.5	4	4.5	6	3.5	3	1.5	2	3	2.5	5	5	7	5.5	1.	2	5.5	6	7
j	5 3 3 0 0 1	4.5	4.5	5.5	6	6.5	6	2.5	3	3.5	2	3	4.5	5	3	5	3.5	2	0	4.5	5	7
k	6 1 3 0 0 2	8.5	7.5	7.5	10	10.5	6	6.5	5	7.5	6	7	8.5	5	1	3	6.5	6	4	2.5	5	7
l	4 1 1 0 0 1	7.5	9.5	10.5	9	9.5	5	2.5	4	6.5	7	8	7.5	6	4	2	5.5	6	5	5.5	1	6
m	1 1 3 0 1 1	9.5	11.5	12.5	11	11.5	1	4.5	4	3.5	9	10	9.5	4	6	10	7.5	8	7	7.5	7	6
n	4 1 3 0 1 1	6.5	8.5	9.5	8	8.5	4	1.5	3	5.5	6	7	7.5	5	3	7	4.5	5	4	4.5	4	5
ŋ	5 1 3 0 1 1	7.5	7.5	8.5	9	9.5	5	1.5	4	4.5	5	6	7.5	6	2	4	5.5	5	3	3.5	4	4
o	6 5 3 1 0 1	4.5	3.5	2.5	4	4.5	10	6.5	7	3.5	2	3	2.5	9	5	5	3.5	3	4	5.5	3	11
ɔ	5.5 6 3 1 0 1	3	2	2	2.5	3	10.5	7	7.5	3	2.5	1.5	2	9.5	6.5	6.5	4	3.5	4.5	7	8.5	11.5

TABLE 8.2

ENGLISH / FRENCH
PHONEME DISTANCES

As vectorial differences
in distinctive features

F.(ɛ - ɔ)

E.(æ - ɯ)

	æ	ɑ	ɒ	aɪ	aʊ	b	d	ð	ɛ	ɜ	ə	ɔɪ	f	g	h	ɪ	ɪ	ʤ	k	l	m
	4	5.5	6	4	4	1	4.5	3	4	5	5	4	2	6	7	4	4.5	5	0	4.5	1
	6.5	7	7	7	7	1	1	2	5.5	5	6	5.5	2	1	2	4.	4.5	0	1.5	1	1
	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	0	0	.5	1	1.5	0	0	0	0	0	0	1	0	0	1	.5	0	0	0	.5	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	2	1	1

ɜ	5.5	6	3	1	1	1	4	2	3	3.5	4	3.5	2.5	3	6.5	7.5	7.5	5	4.5	5.5	2	6.5			
æ	4.5	5.5	3	1	0	1	2.5	3.5	3.5	2	2.5	7	5.5	6	1.5	2	2	.5	3	7	7	2.5	2	4	4.5	7		
ɑ	4.5	6.5	3	1	0	1	1.5	2.5	0.5	1	1.5	10	6.5	7	0.5	3	2	1.5	9	8	8	3.5	3	5	4.5	8		
ɒ	4.5	6.5	3	.5	1	1	2.	3	3	2.5	3	10.5	7	7.5	3	3.5	3.5	3	9.5	9.5	4.5	4	3.5	5.5	4	10.5		
p	1	1	3	0	0	2	9.5	11.5	12.5	11	11.5	1	4.5	4	2.5	9	10	9.5	2	6	8	7.5	1	7	5.5	7		
r	4	1	4	0	0	1	6.5	2.5	9.5	3	7.5	4	1.5	3	5.5	6	7	6.5	5	3	7	4.5	5	6	4.5	6		
ʌ	6.5	1	4	0	0	1	9	8	8	10.5	11	4.5	3	5.5	6	4.5	7.5	9	4.5	4.5	4.5	7	6.5	8	7.5	10		
ɛ	6	2	2	0	0	1	6.5	5.5	5.5	7	6.5	6	3.5	3	5.5	4	5	6.5	5	1	3	4.5	4	3	4.5	5		
ɔ	4	2	3	0	0	2	5.5	7.5	8.5	7	7.5	5	2.5	2	4.5	5	6	5.5	2	4	4	3.5	4	2	4.5	5		
ʃ	5	2	3	0	0	2	6.5	6.5	7.5	8	8.5	6	2.5	3	5.5	4	5	6.5	3	3	3	4.5	4	2	4.5	5		
ɪ	4	1	3	0	0	2	6.5	8.5	9.5	9	8.5	4	1.5	3	5.5	6	7	6.5	3	3	5	4.5	5	4	3.5	5		
u	6	4	3	1	0	1	5.5	4.5	3.5	5	5.5	9	5.5	6	4.5	3	4	5.5	8	7	7	3.5	3	3	4.5	6		
v	2	2	3	0	0	1	6.5	8.5	9.5	8	8.5	12	8.5	1	5.5	6	7	6.5	1	5	7	4.5	5	4	3.5	6		
w	1	3	3	0	0	1	6.5	8.5	9.5	9	8.5	2	5.5	3	5.5	6	7	6.5	3	7	9	4.5	5	4	7.5	8		
y	4.5	4.5	3	1	0	1	3.5	4.5	4.5	3	3.5	8	4.5	5	2.5	2	3	1.5	7	6	6	4.5	4	3	4.5	6		
ɯ	4	3	3	1	0	1	4.5	6.5	6.5	4	4.5	6	3.5	3	3.5	4	5	2.5	5	5	5	1.5	3	2	3.5	5		
z	4	2	3	0	0	1	4.5	6.5	7.5	6	6.5	4	1.5	1	3.5	4	5	4.5	3	3	5	2.5	3	2	3.5	5		
ʒ	5	2	3	0	0	1	5.5	5.5	6.5	7	7.5	5	1.5	2	4.5	3	4	5.5	4	2	4	3.5	3	1	4.5	6		
							14.5	16.5	17.5	16	16.5	6	8.5	9	13.5	14	15	14.5	9	11	15	13.5	13	10	13.5	17		

TABLE 8.3
ENGLISH / FRENCH
PHONEME DISTANCES

As vectorial differences in distinctive features

F.(a - z)

a	4	7	3	0	0	1
c	6	7	3	0	0	1
ã	5.5	7	3	0	1	1
b	1	1	3	0	0	1
d	1	1	3	0	0	1
e	4	5	3	0	0	1
ɛ	4	6	3	0	0	1
ɛ̃	4	6	3	0	1	1
o	5	5.5	3	1	0	1
f	2	2	3	0	0	2
g	6	1	3	0	0	1
i	4	4	3	0	0	1
j	5	3	3	0	0	1
k	6	1	3	0	0	2
l	4	1	1	0	0	1
m	1	1	3	0	1	1
n	4	1	3	0	1	1
ɲ	5	1	3	0	1	1
o	6	5	3	1	0	1
ɔ	5.5	6	3	1	0	1

S.(n - d7)

n	ŋ	ɔ	ɔ̃	ɔɪ	p	ɹ	s	ʃ	tʃ	t	θ	u	o	ʌ	v	w	z	ʒ	ʃʃ	
4.5	6	6	6	6	1	5	4	5	5	4.5	3	6	5.5	5.5	2	1	4	5	5	
1	1	6	6	6	1.5	3	2	2	2	1.5	2	4	5	6.5	2	3	2	2	2	
3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	
0	0	1	1	3	0	0	0	0	2	0	0	15	0	0	0	0	0	0	2	
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	1	1	1	0	2	1	2	2	2	2	2	1	1	1	1	1	1	1	1	

a	7.5	9	4	4.5	4.5	7.5	7	6	7	9	7	7	6.5	4	2	7	7	5	6	7
c	8.5	7	2	3.5	3.5	11.5	7	8	7	9	8	9	7.5	3	1	9	9	7	6	8
ã	7	6.5	3.5	4	4	12	7.5	8.5	7.5	8.5	8.5	8.5	6	3.5	1.5	7.5	7.5	7.5	6.5	6.5
b	4.5	6	11	11.5	11.5	1.5	8	5	6	7	5	4	2.5	9	10	2	3	4	5	7
d	1.5	3	8	8.5	8.5	7.5	5	2	3	5	2	3	6.5	6	7	3	5	1	2	4
e	5.5	7	4	4.5	4.5	7.5	5	4	5	7	5	5	4.5	2	3	5	5	3	4	6
ɛ	6.5	8	3	3.5	3.5	8.5	6	5	6	8	6	6	5.5	3	2	6	6	4	5	7
ɛ̃	5.5	7	4	4.5	4.5	8.5	7	6	7	9	7	7	6.5	4	3	7	7	5	6	8
o	7	7.5	1.5	2	2	10	5.5	6.5	5.5	5.5	6.5	7.5	3	1.5	6.5	7.5	7.5	5.5	6.5	6.5
f	5.5	7	10	10.5	10.5	1.5	7	8	8	5	8	1	2.5	8	9	1	3	3	4	2
g	2.5	1	6	6.5	6.5	6.5	5	4	3	5	3	5	4.5	5	6	5	7	5	2	4
i	4.5	4	5	5.5	5.5	6.5	4	3	4	6	4	4	3.5	3	4	4	4	2	3	5
j	3.5	4	5	5.5	5.5	6.5	3	2	2	4	3	4	3.5	3	4	4	4	2	1	3
k	3.5	2	7	7.5	7.5	5.5	6	3	2	4	2	4	5.5	6	7	6	8	4	3	5
l	3.5	5	10	10.5	10.5	6.5	3	4	5	7	4	5	7.5	8	9	5	7	3	4	6
m	3.5	5	12	12.5	12.5	8.5	9	6	7	9	6	5	6.5	10	11	3	3	5	4	8
n	2.5	2	9	9.5	9.5	8.5	6	3	4	6	3	4	7.5	7	8	4	6	2	3	5
ɲ	2.5	1	8	8.5	8.5	6.5	5	4	3	5	3	5	6.5	6	7	5	7	3	2	4
o	7.5	6	1	1.5	1.5	10.5	6	7	6	6	7	8	5	1	3	8	7	6	5	5
ɔ	8	7.5	5	1	1	11	6.5	7.5	6.5	6.5	7.5	8.5	8	1.5	1.5	9.5	8.5	6.5	7.5	7.5

TABLE 8.4
ENGLISH / FRENCH
PHONEMIC DISTANCES

As vectorial differences
in distinctive features

E.(n - dʒ)

F.(ʒ - ʒ)

	n	ŋ	ɔ	ɔʊ	ɔɪ	p	ɹ	s	ʃ	tʃ	t	θ	ʊ	ɒ	ʌ	v	w	z	ʒ	dʒ
	4.5	6	6	6		1	5	4	5	5	4.5	3	6	5.5	5.5	2	1	4	5	5
	1	1	6	6		1.5	3	2	2	2	1.5	2	4	5	6.5	2	3	2	2	2
	3	3	3	3		3	1	3	3	3	3	3	3	3	3	3	3	3	3	3
	0	0	1	0		0	0	0	0	2	0	0	1.5	.5	0	0	0	0	0	2
	1	1	0	0		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	1	1	1		2	1	2	2	2	2	2	1	1	1	1	1	1	1	1
ɜ	5.5	6	3	1	1	1														
æ	4.5	5.5	3	1	0	1														
ɑ	4.5	6.5	3	1	0	1														
ɑ̃	4.5	6.5	3	.5	1	1														
p	1	1	3	0	0	2														
r	4	1	4	0	0	1														
ə	6.5	1	4	0	0	1														
ɛ	6	2	3	0	0	1														
s	4	2	3	0	0	2														
ʃ	5	2	3	0	0	2														
t	4	1	3	0	0	2														
ɛ	6	4	3	1	0	1														
v	2	2	3	0	0	1														
w	1	3	3	0	0	1														
y	4.5	4.5	3	1	0	1														
ɥ	4	3	3	1	0	1														
z	4	2	3	0	0	1														
ʒ	5	2	3	0	0	1														
∆	0	0	0	1	0	0														

	n	ŋ	ɔ	ɔʊ	ɔɪ	p	ɹ	s	ʃ	tʃ	t	θ	ʊ	ɒ	ʌ	v	w	z	ʒ	dʒ	
ɜ	7	6.5	15	2	2	12	7.5	8.5	7.5	7.5	8.5	7.5	4	2.5	2.5	7.5	7.5	6.5	6.5	17	
æ	6.5	8	2	2.5	2.5	9.5	6	6	6	6	6	7	3.5	2	3	7	7	5	5	15	
ɑ	7.5	9	2	2.5	2.5	10.5	7	7	7	7	7	8	4.5	3	2	8	8	6	6	14	
ɑ̃	6	7.5	3.5	4	4	11	7.5	7.5	7.5	8.5	7.5	8.5	6	3.5	4.5	6.5	7.5	6.5	6.5	16.5	
p	5.5	7	12	12.5	12.5	5	9	4	5	7	4	5	10	10	11	3	5	5	6	8	7
r	2.5	4	9	9.5	9.5	5.5	6	3	4	6	5	4	7.5	7	8	4	6	2	3	5	12
ə	4	2.5	7.5	8	8	6	4.5	5.5	4.5	4.5	4.5	4.5	6	6.5	7.5	4.5	5.5	4.5	5.5	5.5	14.5
ɛ	3.5	2	5	5.5	5.5	6.5	4	5	2	4	2	4	3.5	4	5	4	6	2	1	5	12
s	3.5	5	3	3.5	3.5	3.5	5	0	1	3	1	1	4.5	6	7	3	5	1	2	1	11
ʃ	3.5	4	7	7.5	7.5	4.5	4	1	0	2	1	2	5.5	5	6	4	6	2	1	3	12
t	2.5	4	7	9.5	9.5	3.5	6	1	2	4	1	2	7.5	7	8	4	6	2	3	5	12
ɛ	6.5	5	2	2.5	2.5	7.5	5	6	5	5	6	7	5	9	10	7	7	5	4	4	15
v	4.5	6	9	9.5	9.5	2.5	6	3	4	6	4	2	7.5	7	8	0	2	2	3	5	9
w	6.5	8	9	9.5	9.5	3.5	6	5	6	8	6	4	7.5	7	8	2	3	4	5	7	8
y	5.5	7	3	3.5	3.5	3.5	5	5	5	5	5	6	2.5	2	4	6	6	4	4	7	14
ɥ	4.5	6	5	5.5	5.5	6.5	4	3	4	4	4	4	3.5	4	6	4	4	2	3	9	12
z	2.5	4	7	7.5	7.5	4.5	4	1	2	4	2	2	5.5	5	6	2	4	0	1	8	13
ʒ	2.5	3	6	6.5	6.5	5.5	3	2	1	3	3	2	4.5	4	5	3	5	1	0	8	11
∆	14.5	12	17	17.5	17.5	7.5	10	11	12	14	11	10	15.5	15	16	9	9	10	11	13	

73

Since there is rarely a one-to-one correspondence between the phonemes of different languages, a phoneme in Language A may have to be compared with two or more phonemes of Language X. For example, the Spanish [β] may be compared with the English /b/, /v/, and /w/. In comparing two grammatical or lexical forms the individual phonemes may differ according to position or adjacent sound, as do the two /d/s in the Spanish word /dedo/.

Neither graphic nor orthographic representation of the formal elements of two languages gives a complete picture of the degree of similarity between two forms. A phoneme to phoneme comparison of interfaced equivalents, however, measuring differences in distinctive features reveals similarities in form that may seem quite different to the eye. Whether the letter c, for example, is compared to k, q, or g,- or any other letter, it is always indicated as a difference. The similarities of words written in different orthographic systems can be reduced by transcribing the words into graphic representations of the phonemes so that differences in spelling do not hide similarities in word-form. For example, compare the following:

E. c a t's	E. / k a t s /
G. K a t z e n	G. / k a t s e n /
E. q u a r t e r	E. / k w ɔ r t ə r/
S. c u a r t o	S. / k w a r t o /

The choice of transcription will depend on whether it is the distance in the forms of the language as forms, that is measured, or whether such communication variables as homographs and homophones are accounted for. If distance between spoken forms is measured, a transcription of the phonemes may not be sufficient, since it does not reveal similarities in such audible features as voicing, point and manner of articulation, so that the similarities in such pairs as /p,b/, /t,d/, and /k,g/ are not taken into account. In order to do so,

comparisons based on distinctive feature differences can be used. To illustrate the three possibilities, take one of the above examples of the English/Spanish cognates cotton/algodón:

Othographic	Phonological	Allophonic
c o t t o n	k o t o n	k o t o n
a l g o d o n	a l g o d o n	a l g o d o n
1+1+2+0+1+2+0+0	1+1+2+0+2+2+0	10+10+1 +0 +1+1.5+1.5
D= 7/13 = .54	D = 8/12 = .67	D = 24/120 = .20

After having taken the syllabic distance and accounting for the difference in stress, use of distinctive features of allophonic comparisons make the two words sound much closer than they would if measured from their orthographic or phonological forms. Note that on the same scale of a maximum distinctive difference of 20 between any two sounds of English and Spanish, the distance between the /k/ of English and the /g/ of Spanish is twenty times less in distinctive features than it is in simple graphic representation, where the phonemes are regarded as being entirely different.

Such vectorial tables of distinctive feature differences can be used not only in measuring the difference between the formal elements of the codes, but also in calculating the distances between equivalent stretches of discourse.

Differences or identities in codes do not, as we have seen, necessarily produce corresponding differences in their equivalent messages. The full extent of differences between languages can be seen by measuring how they differ in performance, that is, by observing the codes in action. In other words, measures for distance in discourse must be obtained.

3.3 Limitations of Intercodal Measurement

Differences in codes are no guarantee of corresponding differences in the messages they produce.

The reasons are that the distribution of similarities and differences produced by even the most exhaustive juxtaposition of two codes will not necessarily be identical to those in two equivalent texts. Examine the following and compare the similarities within the first set (Set I) of equivalent sentences as against those within the second set (Set II):

I	II
E. Here comes the old man who is blind; give him your hand.	E. A permanent committee of the government is responsible for changes in the Constitution.
G. Hier kommt der alte Mann der blind ist; gib ihm die Hand.	G. Ein ständiger Regierungsausschuss ist für die Änderungen des Staatsgrundgesetzes verantwortlich.
F. Voici le vieillard qui est aveugle; donne-lui ta main.	F. Un comité permanent du gouvernement est responsable des changements dans la Constitution.

It is obvious at a glance that in one case (I), the English text is much closer to the German than it is to the French, whereas in the second case (II), the converse is true, the English text being closer to the French.

Two languages may have a number of such similarities at one level and not at another. There may be similarities in the vocabulary but not in the grammar; similarities in the grammar but not in the vocabulary; similarities in certain features of pronunciation but not in others. For example, in most respects, French and Hungarian are quite different; but in certain features of pronunciation - phonemes (rounded front vowels) and prosodemes (syllabic equality) they have much in common. Likewise, English and Chinese are described as being highly different languages, yet in some respects English may be more similar to Chinese than it is to German. According to soundings made by

Jespersen, it would seem that the regularity of word-order (SVO) in Modern English prose would range roughly between almost 90% and 97%, as compared with Chinese, which is close to 100%. Contrast this to German where the regularity is barely above 30%.

For practical purposes, the similarity of two languages will be seen in the messages or texts which the languages produce. Since students of language learning and bilingualism are faced with specific cases, it is important to discover techniques for dealing with them. One way is to analyse and measure not only the differences between the codes, but also the distances between the messages or samples of discourse which the codes generate within comparable contexts.

In order to be comparable, two texts must be equivalent. A greeting is comparable to a greeting and not with something else. Any two stretches of speech which convey the same message - a sequence of sentences, a phrase, a word, or a morpheme - can be juxtaposed to an equivalent in the other language. One can even imagine two languages as two equivalent, endless and interfacing texts.

The operations necessary for bringing about this equivalence are functions of the degree of difference - the distance - between the two languages, and can therefore be used as a measure of this distance. These operations range from the simple use of homonyms (homophones and homographs), which are the same in both languages, to complex changes in structure and level.

It is therefore possible to establish a scale which would reflect this difference. If we postulate equivalence in meaning we measure differences in form and function. We can start either from the orthographic form or from phonetically transcribed samples of speech. For purposes of consistency and simplicity, we shall start with the orthographic texts.

The equivalence of two such interfacing texts poses two problems. Within each language, there is a difference between what must be said and what may be said. In some languages there may be only one way of expressing a given message; in others, there may be many ways of uttering the equivalent message. In one language there may be more freedom than in the other. And this difference in degree of freedom is itself a measure of the distance between languages. This degree of freedom is greatly restricted once the stretch of discourse has to function in a context of situation (See 3.2 above).

To be able to compare any two equivalent stretches of speech, however, and measure the difference between them, one must first know the features in which one text can differ from another.

In measuring the difference between two languages in semantically equivalent texts, one can either begin by analysing and classifying their differences, giving values to each type, that is, by taxonomic measurement, or one can arrive at a single measure of the texts as they appear face to face, that is, by integral measurement.

In the first case, taxonomic measurement, one may have to go beyond what meets the eye and relate the text to the code which produced it; in the second case, one measures only what is overtly expressed and is directly observable. Let us begin by analysing the first case, the taxonomic distance in discourse.

4

TAXONOMIC DISTANCES IN DISCOURSE

4.1 Dimensions of Difference

4.1.1 A Typology of Dimensional Differences

4.1.2 An Algebra of Dimensional Differences

4.1.2.1 Notation of Hierarchy
4.1.2.2 Notation of Order
4.1.2.3 Notation of Length
4.1.2.4 Notation of Depth
4.1.2.5 Notation of Elements
4.1.2.6 Notation of Relation
4.1.2.7 Notation of Set

4.1.3 Differential Analysis

4.1.3.1 Number of Dimensions
4.1.3.2 Components of Differences

4.2 Degrees of Difference

4.2.1 Conversion of Hierarchy

4.2.2 Conversion of Order

4.2.3 Conversion of Length

4.2.4 Conversion of Depth

4.2.5 Conversion of Elements

4.2.6 Conversion of Relation

4.2.7 Conversion of Set

4.2.8 Multidimensional Conversion Distance

4. TAXONOMIC DISTANCES IN DISCOURSE

Equivalent stretches of speech can differ in several ways, since all languages have various means of expression, such as intonation, vocabulary, and word-order. Within each of these areas or dimensions there are great differences between languages -- both in type and in degree. In taxonomic measurement, differences of type within these dimensions of interlingual difference must precede the measurement of the degrees of difference.

In order to find measures for different types of difference, it is first necessary to determine what these types are. One must begin by asking a basic question: How can two languages differ in the texts they produce? The dimensions of difference must be determined before the degree of difference.

4.1 Dimensions of Difference

To examine the dimensions of interlingual difference in discourse, the different types will first be isolated and an efficient way developed of referring to them. In other words, before one can engage in any sort of taxonomic measurement, both a typology and an algebra of language differences are required.

4.1.1 A Typology of Dimensional Differences

How can languages differ in form? They should be able to differ in the dimensions common to all languages. What are these dimensions? It seems evident, as we have seen, that all languages, by their very nature must have some sort of elements (sounds, words, sentences) some of which are more inclusive than others, thus forming a hierarchy of inclusion. These elements must appear in some order in speech, one before the other. These sequences of elements must necessarily have length, some being longer than others. The elements, belonging as they do to classes, must also have depth, some permitting more substitutions than others. These elements are composed

of units which are visually or acoustically perceptible, as being identical or equivalent. Some of these elements are bound to have relations with other elements. And finally, all elements belong to sets, the members of which have similar functions.

Two equivalent language texts may differ in any or all of these dimensions. They may differ in the number of echelons in their hierarchies, as might be the case between a highly agglutinative language like Eskimo and a highly analytical one like English. They may differ in the required and optional order of their elements, functioning as units within a structure, as in the case of word order in French, German, English and Latin. They may differ in their length, as measured by the number of elements in the corresponding units, as would be evident from a comparison of Hungarian and Chinese words. Some elements, however, may not even have an equivalent in the other language; others may be left unstated in discourse, so that their length is zero. Others may have an equivalent with something added, as when a verb in one language is equivalent to a verb and a phrase in the other. In some cases, this counterpart, belonging to a different rank in the hierarchy - as when a pronoun in one language has to be rendered by a group of words in the other language - entails by that very fact a difference in length. Equivalent texts may also differ in their depth as measured by the number of alternative units which can replace each unit in the utterance; the pronominal units of Japanese or Bengali give more alternatives than do Dutch or English. Depth is a measure of what may be considered as equivalent. Most obviously, languages differ in the configurations of the elements themselves - the shape of their morphemes (words and word-endings); yet language contact and genetic relationship are the causes of many identities and similarities which tend to reduce this sort of distance between languages - visual identities (homographs like nation

in English and French) and auditory similarities (homophones like /haus/ in English and German). Equivalent texts of two languages may also differ in the number of specified relations between the units, some languages, like German, requiring changes such as the form of the article and the adjective to agree with that of the noun; other languages, like English requiring no such formal relations. Finally the two languages may differ in how they organize their elements into such distinctive sets as form-classes, as when something expressed by a verb in French must appear as an adjective in English. The formal organization of different languages, therefore, may differ in any or all of seven dimensions - hierarchy (H), order (O), length (L), depth (D), element equivalence (E), relationships (R), and set membership (S).

The first taxonomic operation is establishing which part or parts of two semantically equivalent stretches of discourse are equivalent in meaning. This can be done by interfacing the texts and segmenting them into equivalent units. For example:

F.	Il	se lave	le matin./
E.	He	washes	in the morning.

When two equivalent texts have been so segmented so that the equivalent units are comparable, each pair of such units can be examined for the above types of differences.

Equivalent units in some pairs of languages will be found to differ in only one dimension; in some pairs they may differ in two, others in three, in four, or in all seven dimensions. For example, the English and German personal pronouns, although they have the same number (depth), differ in their required relationship, whereas in English and French they differ in both depth and relation; in one case we have a one dimensional difference, in the other, a two dimensional one. For each feature of any pair of

languages being thus compared or contrasted, there are many possibilities, ranging from a single difference in hierarchy (H) to seven simultaneous differences (HOLDERS) (See Table 9).

Below are examples of differences in one, two, three, four, five, six, and all of the seven dimensions and also examples of a number of selected combinations of these components. So as to be able to identify behind the particular words and word groups, the type of arrangements and relationships involved, some way of noting essential differences has to be worked out. For differences within each of the dimensions are not all of the same type or extent. In order to deal with them in specific cases, some sort of simple but adequate notation has to be developed - indeed a sort of algebra, in the sense of a generalized systematic use of symbols to express and analyse the relationship between concepts.

4.1.2 An Algebra of Dimensional Differences

In devising a simple notation for the analysis of language differences in each of the above dimensions, it is important to keep in mind the existence of symbols already in use in both linguistics and mathematics, while attempting to solve conflicts with the maximum of simplicity and elegance. This can be done with the help of a simple model of two different but equivalent texts (See Figure 2).

If the two languages are symbolized as Language A and Language X, the units of their texts as ABC and XYZ, respectively, the constituent elements of each of these units can be /abc/ and /xyz/, so that Unit A is equivalent (≂) to Unit X, B to Y, and C to Z (See Table 10).

With this basic notation we are able to express differences in hierarchy, order, length, depth, element equivalents or identity, relationship, and set membership of two equivalent texts can be expressed with supplementary notation of variables peculiar to certain dimensions.

TABLE 9

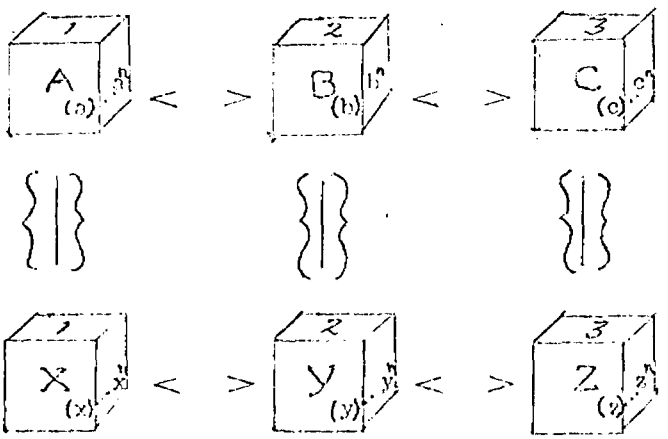
DIMENSIONS OF LANGUAGE DIFFERENCE

1-D	2-D	3-D	4-D	5-D	6-D	7-D
H	HL HO	HLR ESO ERS	HOLD SEOR SODR	ODERS LDERS	OLDERS	
O	OD OR	HOL LSE LDS	HORL SEDR SOLD	LERSH DERSH	LDERSH	
L	OL DR	HOD DSE HDS	HODR SELD SHLD			
D	LR LD	HOR RSE HLE	OLDR SEHD HLRE	DRSHO ERSHO	DERSHO	
E	HR HD	LDR HLS ODE	HLDR SHLR HODE			HOLDERS
R	HE HS	ODR ODS OLE	SEHL SHOD LDRE	ESHOL RSHOL	ERSHOL	
S	OE OS	OLR OLS LRE	SEOD SLDR OLRE	RHOLD SHOLD	RSHOLD	
	LE LS	OLD LRS HRE	SEOL SOLR RHDE	SOLDE HOLDE		
	DE DS	RHD HRS HOE	SELR SRHD HOLE		SHOLDE	
	RE PS	HLD HOS ORE	SEHR SHOL HORE	HLDER OLDER	HOLDER	
	SE	HES ORS DRE	SEHO SHOR ODRE			
		LDE HDE	OLDE HLDE			

Types of differences in: Hierarchy (H), Order (O), Length (L), Depth (D), Elements (E), Relation (R), Set(S).

Figure 2

DIFFERENTIAL MODEL



Algebraic model of equivalent texts in Languages A and X, showing hierarchy (Aa), order (ABC), length (1-2-3), depth (a^n), elements (a,b,c), relations <A,B> and set membership {A,X} .

Example:	A	B	C
	une	grande	table
	|	|	|
	a	big	table
	X	Y	Z

Equivalence (=⌔=):

A =⌔= X , B =⌔= Y , C =⌔= Z.

Congruence (≅):

A ≅ X , B ≅ Y , C ≅ Z.

Identity (≡):

C ≡ Z

Difference (∼):

(ABC) ∼ (XYZ) = A ≇ X) + B ≇ Y)

TABLE 10

SIGNS AND SYMBOLS FOR TAXONOMIC MEASUREMENT

RANKS	UNITS	CATEGORIES	GROUPINGS
A_o sentence	A X	N nominal n pronoun	(a,a) constituents
A: clause	B Y	V verbal v aux. verb	
A; phrase	C Z	D adjectival d determiner	<a,b> relations
A, group		A adverbial a str. adverb	
	ELEMENTS		
A. word		c conjunction p preposition	
-A morpheme	a x		{a,b} sets
≈ A prosodeme	b y	RELATIONS	
/A/ phoneme	c z	Equivalence =◇=	}a,b{ non-sets
A] allophone		Congruence ≅	
		Identity ≡	
		Equality =	
		Difference ∼	

4.1.2.1 Notation of Hierarchy

If hierarchy is further symbolized by quasi-conventional sub-scripts, a limited number of markings is capable of describing any equivalence or difference in rank, from sentence to allophone. These are: sentence (A_o), clause (A:), phrase (A;), word group (A,), word (A.), morpheme (-A), prosodeme ($\bar{A}$), phoneme (/A/), and allophone ([A]).

For example, in the following equivalent texts, one segment, a French phrase (A;), n'importe qui, is equivalent in meaning to an English falling-rising (˅) intonation pattern ($\bar{X}$):

F. Ils n'invitent pas n'importe qui /
E. They don't invite ˅anybody.

A, =<= $\bar{X}$

4.1.2.2 Notation of Order

If A is equivalent in meaning to X, B to Y, and C to Z, identity or difference in order can be indicated simply by the sequence of the letters. For example, the French word répond (A.) is equivalent to the English word answers (X.) and F. toujours (B.) to E. always (Y.) in the following sentences:

F. Il répond toujours /
E. He always answers

A.B. =<= Y.X.

4.1.2.3 Notation of Length

Length is indicated simply by the number of units. To avoid confusion with differences in hierarchy, however, it is useful to stay within the boundaries of the rank being n ed, since the higher the rank, the longer the sequence of units is likely to be. But not necessarily: Thanks. can be just as much a sentence as I thank you very much. In the following example, a French verb comprising of a single word is expressed in English in words:

F. Chercher /

E. Look for.

A. =<>= X.Y.

Length can also be indicated by numbering each component. (See below.)

4.1.2.4 Notation of Depth

Depth can be considered as the number of alternative equivalents which the code permits - not only the number of different vocabulary equivalents, but also the number of phrase, clause, and sentence equivalents; in sum, the number of different ways of establishing an equivalent text. In the example below, the French word lueur can be rendered in two ways in English, by glow or gleam:

F. Lueur /

E. Glow
Gleam

A. =<> X.
Y.

Depth may also be indicated by noting the number of alternatives as an index in superscript. This is especially useful in measuring degrees of differences.

$A. = X^2.$

4.1.2.5 Notation of Elements

In noting the similarities and differences of the constituent elements of two equivalent texts, a lower-case letter can be used for the unit or segment being analysed - (a) for A, (x) for X. For example, in the French word group (A,) pauvres parents, the first word pauvres can be noted as the first constituent (a_1) and the second word, parents as the second constituent (a_2), numbering from left to right. Comparison is also done from left to right, so that if on the right side of the equation an element appears which is identical to its equivalent on the left side it can be so indicated, as is the orthographic identity of the word parents in both texts:

F. pauvres parents /

E. poor parents

A.B.=<>= X.B.

or

$A(a_1a_2) =<>= X(x_1a_2)$

4.1.2.6 Notation of Relations

Any relationship that needs indicating can be expressed between the conventional relationship brackets $< >$. But only significant differences need be marked. For example:

F. grande maison /

E. big house. $\quad <A.B> =\!\!> X.Y.$

Many types of concord and government can thus be stated with varying degrees of refinement.

4.1.2.7 Notation of Set

If two units or elements are not members of the same or equivalent set (form-classes, parts of speech, grammatical category, phrase type, etc.) this can be shown by the conventional set-builder brace $\{\ \}$. To indicate that they belong to different sets, the type of category may be indexed, (e.g.N for nominal, V for verbal) $\{A_c\} =\!\!> \{X_c\}_v$

or invert the braces : for example, F. Attention/E.Look out. $\}A.\{ =\!\!> \}X.Y.\{$

4.1.3 Differential Analysis

An equivalent pair may differ in only one dimension, two dimensions, three and more dimensions. The complexity of the conversion difference may be reflected in the complexity of the formula.

The above notation is sufficient to indicate differences in any combination of dimensions; as the following examples indicated, this can be done both for the description of the number of differing dimensions and for the inclusion of various dimensional components.

4.1.3.1 Number of dimensions

1-D Differences

F. Oui, effectivement je suis adulte./

E. Yes, I 'am an adult.

A. effectivement / ' $\bar{X}$

H (A/X): $A. =\!\!> \bar{X}$

2-D Differences

F. Ce fut le plus grand que l'on ait jamais vu/

E. It was the largest ever seen.

A ($(a_1a_2a_3a_4a_5)$:	que l'on ait jamais	/	ever	X(x).
B(b).	vu	/	seen	Y(y).

HE (A/X): A $(a_1a_2a_3a_4a_5)$: B(b) $\eqsim$ X(x). Y(y).

3-D Differences

A (a).	G. Bitte (gesturing)	/	E. Have a seat	X $(x_1x_2x_3)$.
B (b).	Danke.	/	Thanks	Y (y).
C (c).	Bitte	/	(nil)	

HEL (A/X): A(a). B(b). C (c). $\eqsim$ X $(x_1x_2x_3)$. Y(y).

4-D Differences

F. Son frère a le nez long./

E. His brother has a long, protruding nose.

A $\{a.\}_{d_1}$	le	/	a	X $\{x.\}_{d_1}$
B<b a>	nez	/	nose	Y (y)
C (c).	long	/	long, protruding,	Z (c) Z (z)

DOSE (A/X): A $\{a.\}_{d_1}$ B<b a> C(c) $\eqsim$ X $\{x.\}_{d_2}$ $^{Z(c).}_{Z(z).}$ $Y_{(y)}$

or: $\{A.\}_1$ B<a>. C $\eqsim$ $\{X.\}_1$ $^{Z(c).}_{Z(z).}$ Y.

5-D Differences

F. Le chien était en train d'enlever avec la langue du chocolat qui se trouvait sur sa figure

E. The dog was licking chocolate off his face.

A;	était en train de	/	was ....-ing	X (x.-y)
$B(b;)(c,)\left\{b:\right\}_v$	d'enlever avec la langue qui se trouvait sur	/ /	lick ... off	Y (y.)z. $\left\{y\right\}_a$
C,	du chocolat	/	chocolate	Z.

HOLES (A/X): A; $B(b;)c,\left\{b:\right\}_v$ ≃ X (x.-y) Y(y.)z. $\left\{y:\right\}_a$

6-D Differences

F. On dit que le roi Abdul est longtemps parti./

E. King Abdul has alledgedly left long ago.

A:	On dit que	/	alledgedly	X.
B $(b_1b_2b_3)$,	le roi Abdul	/	King Abdul	$Y(y\ b_3)$,
C (c_1) $\left\{c\right\}_A<c_3b_1>$;	est longement parti	/ /	left long ago	$Z(z)\left\{z,\right\}_D$;

or

$C(c_1).\left\{c_2z_2\right\}.\langle c_3, c_1\ b_1\rangle.$

or

$C\ c_1\ \ c_2\ \ c_3\ \langle c_1\ b_1\rangle.$ $\qquad Z\ z_1.\left\{z_2,\ c_2\right\},$

HOLERS (A/X): $A{:}B\ (b_1b_2b_3), Cc_1\ c_2\ c_3<c_1b>$; ≃ X. Y $(y\ b_3)$, Z $z_1\left\{z_2c_2\right\}$,

7-D Differences

F. Le professeur Martin aurait été congédié sans cérémonie.

E. It seems that Professor Martin has been summarily fired (let off).

$A(a_1a_2a_3)$	Le professeur Martin	/ Professor Martin	X $(x\ a_3)$
B	-ait	/ it seems that	Y:

$C(c_1 b) :: c_2 \langle c_3 a_1 \rangle \left\{ C_x \right\}_N$ aura été congédié / has been summarily sans cérémonie fired (let off) $Z\left(z, \left\{ z \right\}_n z^2\right)$

HOLDERS (A/X): $A\ (a_1 a_2 a_3)\ C\ (c_1 b) :: C_2 \langle c_3 a_1 \rangle \left\{ C_x \right\}_N \Rightarrow Y$; $X\ (x=3)$, $Z\left(z, \left\{ z \right\}_n z^2\right)$

4.1.3.2 Components of Differences

In each of these groups (1-7 dimensions) which have just been exemplified, the actual dimensional components may be quite different. For example, a two-dimensional difference may be of hierarchy and order (HO), length and relations (LR) or any of the other 19 combinations of two-dimensional components (See Table 9.) Here are a few examples of such combinations:

L $\bar{A} \Rightarrow \bar{A}\ \bar{B}$ F. Mayonnaise/ J.Mayonezu F. (s) = J. (ss)

O $A.B. \Rightarrow B.A.$ F. crédit social / E. social credit

SE $\left\{ A. \right\}_{d_1} \Rightarrow \left\{ X. \right\}_{d_2}$ F. Levez la main./ E. Lift up your hand.

HE $A; \Rightarrow X_o$ F. Un vrai sauvage/ E. He's just a barbarian.

SHE $\left\{ A, \right\}_N B.C; \Rightarrow \left\{ \quad \right\}_m YZ.$ F. Le temps est a la pluie. / E. It looks like rain.

LEO $A,B,A, \Rightarrow X.Y.$ F. Il se rend chez-lui en voiture./ E. He drives home.

LES $A.B; \Rightarrow X.$ F. Il faut le frapper a grands coups./ E. You've got to thump it.

DES $\left\{ \begin{matrix} A. \\ B. \end{matrix} \right\}_o \Rightarrow \left\{ X. \right\}_n$ F. Il se rendit fier,orgueilleux./ E.He surrendered proudly.

REL $A,B. \ldots \Rightarrow X,Y.Z \langle X.z \rangle$: F. Il n'était pas prêt,n'est-ce pas? / E. He wasn't ready, was he?

LEOS $\left\{ A. \right\}_v \left\{ B. \right\}_N \Rightarrow \left\{ Y. \right\}_v \left\{ X. \right\}_1$ F. Il monta d'un pas lourd l'escalier./

or:

$A. \left\{ B. \right\}_N \Rightarrow Y. \left\{ X \right\}_1$ E. He plodded up the stairs.

4.2 Degrees of Difference

Differences between languages are not only matters of type; they are also matters of degree. Differences in length, for example, vary according to the number of units by which one stretch of speech in one language exceeds its equivalent in the other language.

In problems of measurement, numbers as well as symbols are needed. These can be incorporated as indices of constituent elements - $(a_1a_2a_3a_4e_n)$. Using the simple procedures of ordinary algebra, both positive and negative numbers can be included, each side of the equation standing for a stretch of discourse in one of the languages: $(A = A)$ and $(A - A = 0)$.

Either the conversion distance or the differential distance between two equivalent stretches of speech can thus be measured. In measuring the conversion distance, it is essential to begin by establishing exactly what is being converted into what, since the operation in one direction may be more complex that it would be in the other. One might imagine that the point of departure, or source language (S) would be on the left of the equation, and the point of arrival, or target language (T) would be on the right. But such is not the case, since the difference must represent the number of additional distinctions to be made by the speaker of the source language in order to equal those of the target language. The formula for measuring the taxonomic conversion distance in discourse must therefore be: $D(S/T) = \left\{T\right\} - \left\{S\right\}$ To simplify the computation, while increasing the range of possible sequences, one set of letters is sufficient beginning with the first letter of the alphabet.

$$\left\{T\right\} - \left\{S\right\} = T\ (A\ ) - S\ (A.) = A. - A. = 0$$

Thus, with a slight modification in the notation, degrees of difference between two equivalent texts can be measured in each of seven dimensions. The texts must first be processed to determine which segments are equivalent.

Here is an example of the procedure:

1. Determine direction of the conversion, that is, which is the target language (T) and which is the source language (S). Write target text before (left or above) source text (right or below).

2. Segment texts for semantic equivalence:

   F. On dit/ qu'/ il/ est parti/ . = 5 segments

   E. He/ has/ alledgedly/ left. = 5 segments

   total segments 10

3. Label segments in target language sequentially (ABC..) from left to right.

   A B C

   On dit qu'/ il / est parti./

4. Give same label to corresponding segment in the source language.

   He / has / alledgedly / left./
   B C A C

5 Establish basic equation: ABC = BCAC

6. Work out conversion equation as indicated below.

7. Express results as powers of each dimension of difference;

   n n n n n n n
   H O L D E R S .

In this way, conversion distances can be separately calculated for differences in hierarchy, order, length, element equivalence, relations and set membership. Or, expressed as a formula:

$$D(S/T) = H(T-S)\ O(T-S)\ L(T-S)\ D(T-S)\ E(T-S)\ R(T-S)\ S(T-S)$$

4.2.1 Conversion of Hierarchy: H (T-S)

If any element in a sentence must change its hierarchical rank

so as to be made equivalent to an element in another language, the number of ranks or steps it has to travel up or down the hierarchical ladder may be used as a measure of the conversion distance in hierarchy. The numerical values attributable to each rank or rung of the ladder must have some relation to its degree of inclusiveness. The highest number is therefore attributed to the highest rank, the text, or for convenience, the sentence ($A_{\circ}$) and the lowest number to the least inclusive, the allophone or allophonic combination ($\dot{A}$). If we operate with nine ranks, the sentence ($A_{\circ}$) appears as Rank 9, the clause (A:) as Rank 8, the phrase (A;) as Rank 7, the group (A,) as Rank 6, the word (A.) as Rank 5, the morpheme (-A) as Rank 4, the prosodeme ($\bar{A}$) as Rank 3, the phoneme (/A/) as Rank 2, and the allophone or speech sound ([A]) as Rank 1 at the bottom of the ladder. The number of steps one must travel up (+value) or down (-value) the ladder to convert a stretch of speech from one language to another in this way represents the distance in hierarchical conversion.

For example, to make the intonation pattern of Anytime in the source language (S), equivalent to a stretch of speech in the target language (T), i.e. à toute heure, we calculate how many rungs higher is the target from the source, that is, T -S. Compare:

	Target Lg. (T)	-	Source Lg. (S)	
	9 A		A 9	
	8 A:		A: 8	
Je ne suis pas libre à toute heure	7 A;		A; 7	
	6 A,		A, 6	
	5 A.		A. 5	
	4 -A		-A 4	
	3 $\bar{A}$		$\bar{A}$ 3	I'm not free Anytime.=A; - $\bar{A}$
	2 /A/		/A/ 2	
	1 A		A 1	

If (T - S) = $\bar{A}$; - A = 7A - 3A = 4

By prefixing the corresponding rank values to each stretch of speech being measured, we can simply subtract one from the other; as in the above example.

4.2.2 Conversion of Order: O (T-S)

If the order of a stretch of speech in one language must be changed to make it congruent with its equivalent in another language, the extent of such a change may be used as a measure of the conversion distance to be travelled. The extent of the conversion can be measured by the number of changes in position that have to be made, and the difference in the position that equivalents may have in each sequence. These differences may be quantified simply by calculating the number and range of position switches needed to make equivalents congruent. Number the constituent elements in each sequence after interfacing their equivalents. For example, F. observation unique / E. unique observation, would be $A_1B_2 - B_2A_1$ Difference must, of course, appear in positive numbers, since order, being exclusively differential, does not involve questions of more or less. If the measurement is in conversion rather than differential distance, the results can be divided by two. So that, $\frac{A_1B_2}{A_2B_1}$

$$\frac{1+1}{2} = 1$$

The difference may be that of a single item removed from the position of the equivalent one, as the adjective in the above example comes before the noun in one language and after it in the other. Or the distance between the items may encompass most of the words in the sentence. Compare:

G. Er hatte die Festung allein gegen einer Angriff verteidigt. /

E. He had defended the fortress alone against an attack.

Segmenting one text according to the equivalents of the other and by subtracting the order number of the equivalents, gives the following:

	1	2	3	4	5	6	7	8	9
A:	Er	hatte	die	Festung	allein	gegen	einer	Angriff	verteidigt.
X:	He	had	the	fortress	alone	against	an	attack	defended.
	1	2	4	5	6	7	8	9	3

$0 + 0 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 6 = 12$

4.2.3 Conversion of Length: L (T - S)

Anyone who has compared a translation with its original may have been struck by the fact that one of the versions takes more space than the other, or that a single word in the source language has been rendered by a number of words in the target language. Differences in length are measured in terms of immediate constituents, the length of words in syllables, phonemes or graphemes - A (abc). - the length of word groups in words - A (a.b.c.), - and so on. This is not the case for differences in hierarchy, some of which also involve differences in length. The difference may be measured for each of the equivalent segments by counting the immediate constituent units (phonemes or letters, syllables, morphemes, word groups, parases, clauses or sentences) and subtracting one from the other.

A simple way to calculate difference in length is to subtract the values of the final indices of each eqrivalent segment. For example, to convert (F. — E.) French le viellard to English the old man, requires one extra word.

L (E.-F.) = A ($a_1a_2a_3$), - A (a_1a_2), = $a_3 - a_2 = 3 - 2 = 1$.

Converting in the opposite direction: L (F. - E.) = A (a_1a_2),- A ($a_1a_2a_3$), = $2 - 3 = -1$, or one word less.

Differences in length may be contiguous ( ) or non-contiguous ( ). It is therefore necessary to interface both texts for optimal equivalence in order to be able to measure with accuracy any difference in length (See below). Compare, for example, the English (the post) with the French equivalent (les postes) and with the Japanese (posuto).

F. postes = 6	J. posutu = 6
E. post = 4	E. pos t = 4
2	2
L (F.- E.).= 2	L (J.- E.).= 2

Differences in order, already measured by a previous formula, will, in this way, not be confused with differences in length.

In analysing two equivalent segments, however, there is always the possibility of confusion between length and hierarchy. Examine the second word in the Hungarian sentence: "Én hotelban vagyok, nem rokonaimnál." (I'm in a hotel, not with my relatives.) Even if another word like "szálloában" were used in this type of sentence, it would still be a single word. In the German equivalent, however, it could be two words (im Hotel), and in French, three words (dans un hôtel). Compare:

H.	hotelban			= 1 (E. - H.) = 3 - 1 = 2
G.	im		Hotel	= 2 (E. - G.) = 3 - 2 = 1
F.	dans	un	hôtel	= 3 (E. - F.) = 3 - 3 = 0
E.	in	a	hotel	= 3

Therefore, one might imagine that:

L (E. - H.); = 2 L (E. - G.); = 1 L (E. - F.); = 0

There is something fundamentally different, however, between the structure of Hungarian on the one hand and that of German, French and English on the other. The difference of the equivalence is clearly of another dimension - that of hierarchy. For the Hungarian equivalent is expressed within the word-structure of the language. One can either be satisfied with this distinction and express the difference as: H (E. - H.) = 7A - 5A = 2. Or the difference in

length can also be marked: (L (E. - H.) = A ($a_1 a_2 a_3$), - A (a_1) = $a_3 - a_1 = 2$.

Since indices of constituents renumbered sequentially ($a_1 a_2 a_3$) for each language do not permit one to distinguish between the constituents of one language and those of the other, the same number cannot now be used to represent different constituents . If the same indices are to be used for both measurement of length and element equivalence, then the sequence of constituent elements of the source language must follow those of the target language A ($a_1 a_2 a_3$) - A ($a_4 a_5$). Therefore, the sequences must be interfaced from left to right, with a value of one for each constituent (See 4.2.5 below).

4.2.4 Conversion of Depth: D (T - S)

Anyone who has attempted to translate a text from one language into another may have been embarassed by the number of ways the same thing can be expressed in the target language. Some languages may be richer than other in certain types of expressions; they may, as it were, go deeper into the subject.

The difficulty of measuring differences in depth between two equivalent stretches of speech is that the alternatives are seldom expressed. In many cases it is necessary to go beyond the textual evidence and measure alternatives from the code. The more alternatives, the more powerful the language is in this area of expression. Indeed, these alternatives can be conceived as powers to be added and subtracted: the total number of the powers in the source language can be subtracted from the power value of the corresponding segment in the target language. This gives a measure of distance in depth. Examine the following:

F. Je l'ai vu briller/ E. I saw it shine
glow
gleam
glitter
shimmer
flare
beam

Or, $D(E. - F.) = A.^{7} - A.^{1} = 6$

Depth measurement may be applied not only to word differences but indeed to differences in any rank of the hierarchy.

Since conversion distance measures only one direction at a time, one must avoid concluding that the fact of a target language being able to express the same thing in a number of ways means that the source language has only one way of expressing it. In order to find out the difference in the potential of each language, the number of possible ways of saying the same thing in one language would have to be subtracted from the number of ways of saying it in the other (See 3.1 and 3.2).

4.2.5 Conversion of Elements: E (T - S)

People learning a foreign language are sometimes pleasantly surprised to find that a certain word is the same in form and meaning as it is in their mother tongue. A learner of French or English, for example, recognizes such written words as page, pipe and nation as familiar. The presence of such words which are contextually identical in form and meaning may make it easier to convert comprehension habits from one language to the other, since the conversion distance is thereby reduced. But how can this conversion distance be measured? The simplest way would be to subtract the number of identical constituents from the number of different ones, making sure to consider only the immediate constituents of the segments being compared. For example:

F. à son hôtel / E. at his hotel =

$A(a_1a_2a_3), = A(a_4a_5a_3) =$

$$\frac{\begin{matrix} a_1a_2a_3 \\ a_3a_4a_5 \end{matrix}}{1+1+\ 0+1+1 = 4} = E(F. - F.) = 4$$

The value of an identical word in two equivalent segments, however, depends on the length of segments; one in three makes the text more similar than one in six. Since the segments may not be equal in length, the number of their identical interfacing elements can be expressed as a proportion of their combined total. For example:

F. à la page six / E. on page six

$A\ (a_1a_2a_5a_4) = A\ (a_5a_3a_4)$

$a_1a_2a_3a_4 = 4$

$a_3a_4a_5 = 3$

$1 + 1 + 0 + 0 + 1 = 3 \qquad 7 \qquad E\ (F. - E.) = 3/7 = .43$

Since we are calculating in terms of immediate constituents, we take into account only identities at the level of these constituents, e.g. a word is either the same or different. In the last analysis, to take all similarities into account, we must deal in the constituents of constituents, e.g. the letters, graphemes or phonemes of the word. For example:

F. une longue planche / E. a long plank

$A_1B_1C_1 = A_2B_2C_2$

$A_1 - A_2 = A_1\ (a_1a_2a_3) - A_2\ (a_4) = 4/4 = 1$

$B_1 - B_2 = B_1\ (b_1b_2b_3b_4b_5b_6) - B_2\ (b_1b_2b_3b_4) = 2/10 = .20$

$C_1 - C_2 = C_1\ (c_1c_2c_3c_4c_5c_6c_7) - C_2\ (c_1c_2c_3c_4) = 4/12 = .33$

So that $E\ (F. - E.) = F.\ (A_1B_1C_1) - E.\ (A_2B_2C_2) = \frac{1 + .20 + .33}{3} - \frac{1.53}{3} = .51$

4.2.6 Conversion of Relations: R (T - S)

Two equivalent stretches of speech may be identical in hierarchy, order, length, and depth of their constituent elements and yet differ widely in the number of obligatory relations between elements. These constraints may be measured by counting their number in each segment and by subtracting the totals.

	le vieux cheval		the old horse
French.		/ English.	
	la vieille maison		the old house

F.(2) - E.(0) = ?

In other words, in each of the two equivalent 3-word stretches of speech, there are two constraints more in French than there are in English. Or, R (F. - E.) = 2 . With the same notation as above, relations may be indicated thus< > , each being applied to a constraint in the system (See 4.1.12). Degree of difference is obtained by counting the number of such relations on each side of the equation and subtracting the difference. For example:

F. ces vieilles chaussures / E. these old shoes

$A\langle a_1 a_2\rangle\langle a_2 a_3\rangle - A\langle a_4 (a_5) a_6\rangle = A\langle 2\rangle - A\langle 1\rangle = 1$

As with length, relations may be contiguous or non-contiguous; they can therefore be computed in the same fashion (See 4.2.3).

To obtain more precise measurements, the relationships may be categorized; for example, number (1), gender (2), case (3), etc. , so that difference in type of relation can, if desired, be taken into account. Consider the following: F. Ils ont levé la main./ E. They lifted up their hands.

$\langle A\rangle_2 - \langle A\rangle_1 = 1$

Since F. la and E. their do not belong to the same sub-set of determiners, this too would have to be taken into account (See 4.2.7 below).

4.2.7 Conversion of Sets: S (T-S)

If an item belongs to one set or category (e.g. nominal) in one language, its equivalent in another language may belong to a different category. Verbs may sometimes be equivalent to nouns (e.g. E. Rent a car. / F. Location d'autos.) and nouns to verbs (E. Adjournment: 5 p.m./ F. La séance s'est levé à 17 h.). In fact, any item in any language can theoretically belong to a different category or grammatical set when translated into another language.

In measuring this dimension of difference, one may either take into account only the identity or lack of identity in set or try to measure the degree of difference. In the first case, the constituents which do not belong to the same set as their equivalents may be marked by inverted (non-set) braces. For example:

E. Gargle twice a day. / F. Gargarisme deux fois par jour.

$$\text{A.} - \}\text{A.}\{ \quad = \quad \}1\{$$

In the second case, the differences in set membership of equivalent elements may be measured along any appropriate scale of difference between whatever sets each language possesses. For example:

E. He was ski-ing. / F. Il faisait du ski.

$$\{\text{A-}\}_4 \quad - \quad \{\text{A.}\}_1 \quad = \quad 3$$

In French and English, scales may be so refined as to include differences between sub-classes of determiners, or they can be limited to such gross distinctions as nouns (N), adjectives (D), adverbs (A) and verbs (V) and other structure words (S). Compare:

	English		French	
	N 1		1 N	J'ai faim.
I'm hungry	D 2		2 D	
	A 3		3 A	
	V 4		4 V	
	S 5		5 S	

$$\{\text{A.}\}_2 \quad - \quad \{\text{A.}\}_1 \quad = \quad 1$$

4.2.8 Multidimensional Conversion Distance

Values can now be given to each dimension, obtaining a better picture of the extent and degree of difference between two equivalent texts. To make clear how this can be done, here are a few examples:

L F. manicure / J. manikyua s/ss

$\bar{A} = \bar{\bar{A}}\bar{\bar{A}} = L(1\text{-}2) = L^{-1}$

LE F. Ses machoires étaient comme soudées./

E. His jaws were locked.

$A(a_1a_2) - A(a_1) = L(2\text{-}1), E(3) = L^{+1}E^{3}$

F. trotiner / E. toddle along

$A(a_2) - A(a_2a_3) = L(1\text{-}2), E(3) = L^{-1}E^{3}$

SHE F. Elle resonne lugubrement. /

E. It has a sinister sound.

$\{A.\}_{A} - \{A,\}_{N} = S(3\text{-}1), H(3\text{-}6), E(1+2) = S^{-2}H^{-1}E^{3}$

DES F. Il se tortura à computer son impot. /

E. He wracked his brains struggling, straining to figure out his taxes.

$\{A.\}^{1}_{S} - \{A.\}^{2}_{VH} = D(1\text{-}2), E(1+2), S(3\text{-}3) = D^{-1}E^{3}S^{+2}$

IES F. Il se sentit vide de tout. /

E. All he felt was an emptiness.

$A\{a_1a_2a_3\}_{A} - A\{a_1\}_{N} = L(+2), '(4), S(+1) = L^{+2}E^{4}S^{+1}$

LEO E. I like last year's dresses better./

F. J'aime mieux les robes de l'an passé.

$A(a_1).\ B(b_1b_2), C(c_1).\ A(a_1).\ -\ A(a_3a_4), C(c_2c_3), B(b_3b_4b_5b_6);$ =

L(2-2) + L(2-4) + L(1-2),

E (12), 0 (4) = $L^{-3}E^{12}0^4$

LEOS F. Elle cria au meurtre comme une perdue. /

E. She shrieked blue murder.

$A(a_1).B\ (\{b_1\}_s b_2),\ A(a_2a_3a_4);\ --\ A(a_5).B\ (\{b_3\}_{is} b_5),$ =

$L(4\text{-}1),\ E(9),\ 0\binom{123}{121} = 2),\ S(5\text{-}2)$ = $L^{+3}E^{9}0^{2}S^{+3}$

These taxonomic measurements have the advantage of indicating not only the degree of difference between languages as manifested in their production of discourse, but also the dimensions in which they differ. They indicate both the how and the how much. They have the disadvantages, however, of obliging one to calculate as many measures as there are dimensions, each measurement being dependent on correct and accurate classification and interpretation of the categories in the language being compared. One must know something about how the codes have produced the corresponding messages, what other alternatives are possible, whether relations are compulsory or optional, which grammatical classes are being related, and many other such points. In other words, the measurements measure more than meets the eye.

A direct, objective and - for certain purposes - more useful type of measurement could be obtained by limiting the differences to what is immediately observable. It would also be useful for certain purposes, instead of having seven measures of difference between two equivalent texts, to come out with a single figure. These objectives may be reached through the elaboration of integral measurements.

5

INTEGRAL DISTANCES IN DISCOURSE

5.0 Processing the Texts

- 5.0.1 Orthographic Texts
- 5.0.2 Phonetic Texts
  - 5.0.2.1 Syllabic Segmentation
  - 5.0.2.2 Phoneme Segmentation

5.1 Static Measurement

- 5.1.1 Supra-Prosodic Segments
- 5.1.2 Infra-Prosodic Segments
  - 5.1.2.1 Syllabic Units
  - 5.1.2.2 Phoneme Units

5.2 Kinetic Measurement

- 5.2.1 Supra-Prosodic Sequences
- 5.2.2 Infra-Prosodic Sequences
- 5.2.3 Integration

5. INTEGRAL DISTANCES IN DISCOURSE

In contradistinction to the taxonomic measures of each dimension of difference between two equivalent stretches of discourse, is the measurement of formal differences between two running texts. In this second type of measurement, a running text in one language is matched with a corresponding text in another language so as to show the extent of divergence as a single measure.

5.0 Processing the Texts

Before any systematic measurement is made, however, it is useful to process the two texts in such a way as to avoid errors in analysis and computation. The processing involves three main operations - interfacing, segmental matching, and numerical marking.

To begin with, both texts are interfaced, one under the other. The interfaced texts are then studied to find out exactly what corresponds to what. And on this basis the texts are segmented. Segmentation by layers of immediate constituents, as is the common practice in bilingual descriptive analysis, may not always be the best type, since in longer or more complex stretches of discourse, it is likely to prolong the analytical phase of the operation without contributing much to the end product. What is suggested here is a type of segmentation by minimal equivalence. This minimal equivalence represents the smallest segments in each text that can, irrespective of order, be considered as equivalent.

Each segment in one text is then matched by connecting lines to the corresponding segment in the other text, and, if necessary, respaced. Matched segments can be marked and numbered to permit easy identification. For example:

F.	Le patron	a subi	un	mauvais	coup./
E.	The boss	had	a stroke of	bad	luck.
		2	3	4	5

This produces two interfacing sequences of segments of minimal semantic equivalence. The process of segmentation can be repeated until the smallest comparable unit has been reached. The end-point will depend on the delicacy of analysis required. If the grapheme or letter is the ultimate unit, it is likely to show results somewhat different than would the use of the phoneme as the smallest formal unit; but, if one is comparing the languages as to how they look, it might be the more appropriate form. In this case, one would be comparing stretches of discourse in orthographic transcription; in the second case, texts in prosodic, phonological or phonetic transcription would be required. Let us first study the segmentation of orthographic texts and leave until later the study of texts in phonetic transcription.

5.0.1 Orthographic Texts

In treating the processing of texts, the operations are indicated on the left and the corresponding examples on the right.

Operations	Examples
1. Interface two equivalent texts A and X.	A - Le patron a subi un mauvais coup. X - The boss had a stroke of bad luck.
2. Mark equivalent segments by overline (a) and underline (x) indicating equivalence with vertical line a | x	1 2 3 4 5 6 A: Le patron a subi un mauvais coup | | / | | | X: The boss had a stroke of bad luck 1 2 3 4 5 6
3. Number segments for positions (P) left to right, opposite each position, indicating number of elements (E) (e.g. words) in the segment	P(a) 1 2 3 4 5 6 E(a) 1 1 2 1 1 1 E(x) 1 1 1 3 1 1 P(x) 1 2 3 4 5 6

4. Starting with inner series rearrange each interfacing sequence according to lines of equivalence, subtracting the greater number from the lesser, and summing the differences

$E(a)$ 1 1 2 1 1 1

$E(x)$ 1 1 1 3 1 1

$\Sigma E(a-x) =$ 0 +0 +1 +2 +0 +0 = 3

$P(a)$: 1 2 3 4 5 6

$P(x)$: 1 2 3 4 5 6

$\Sigma P(a-x) =$ 0 +0 +0 +0 +0 +0 = 0

5. List results of difference in the number of their elements $E(a-x)$ and their positions in sequence $P(a-x)$.

$E(a-x) =$ 3

$P(a-x) =$ 0

6. Take each equivalent pair in turn for comparison of their formal constituents (i.e. the letters).

Le patron

The boss

7. Segment into interfacing graphemes, and join equivalents

L e p a t r o n

Th e b o s s

8. Total graphemes(e) in each segment and subtract in real numbers the combined differences in length

2 6

$- \underline{3}$ $\underline{4}$

$\Sigma e(a-x) = 1 + 2 + 3$

9. Add all graphemes

2 6

$+ \underline{3}$ $\underline{4}$

5 10

10. Count matched pairs 2 0

11. Add total graphemes in segments: $\Sigma e(a+x)_1 + \Sigma e(a+x)_2$ 5 + 10 = 15

12. Total identical (=)elements(p) $\Sigma e(a=x)$ = 2 + 0 = 2

13. Express totals as a proportion or percentage. 2/15 = .13

3. Re-arrange each interfacing sequence according to lines of equivalence, subtracting the greater number from the lesser, and summing the differences.

$R_{(a)}$: 2 1 1 1

$R_{(x)}$: 1 1 0 0

$R(a-x)$: 1 + 0 + 1 + 1 = 3

ii) Semantic Components

	English	French
1. Count number of dictionary X meanings of A element.	boss	(patron (chef (contremaitre
	$_{M}(a \Rightarrow x)$	= 3
2. Count number of dictionary A meanings of corresponding X element.	French patron	English (boss (master (head (owner (employer (pattern (model
	$_{M}(x \Rightarrow a)$	= 7
3. Subtract equivalence		$_{M}(a \sim x)$ = 7 -3 = 4
4. Repeat calculations for other words in text (e.g. the(30) - le (43) = 13)and express as proportion of total meanings.	The/ le	boss/patron
	le = 43 43 +	7 7
	The - 30 + 30	-3 +3
	$_{M}(a \sim x)$ = 13 / 73	4 /10
5. Interface graphic distances(G)	G = 3/5 +	10/6 etc. = 13
6. Reduce both to percentages	M = .19 +	.04
	G = .6 +	1.0
7. Express as sequence of ratios	.19/.6 +	.04/1 + ...

Such measures of interlingual distance could be experimentally correlated with psycholinguistic measures of interlingual identification.

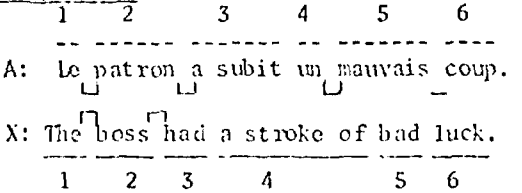
14. Note differences in segmental order, length, word length and word elements.	Seg. order	0/12
	Seg. length	3/16
	Word length	3/16
	Word elements	3/15
15. Express in percentages or equivalent	Seg. order	.0
	Seg. length	.12
	Word length	.12
	Word elements	.13
16. Total and average	37/4	.09

This indicates the average distance in the formal characteristics of the two stretches of discourse. Note that the measurement is based entirely on the evidence supplied by the texts and does not refer to the corresponding codes.

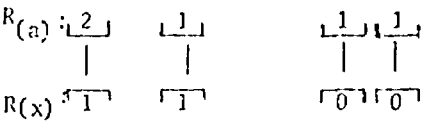
More information can be quantified concerning the functional difference between the texts, especially as they affect the speaker, only by referring to the respective codes. This advantage may be counter-balanced, however, by the fact that the analysis may become less objective and accurate, depending as it does on the completeness of grammatical and semantic knowledge on the operations of the code behind the message. Here, nevertheless, are examples of such expansions designed to show differences in relationships and semantic components.

i) Relacionships (R)

1. Indicate syntactic dependence with horizontal line, thus:



2. Number the compulsory categories of relationship (R), e.g. gender and number.



5.0.2 Phonetic Texts

When the analysis reaches the word-level, there is the question of whether to use orthographic or phonetic texts. The above examples made use of orthographic texts; let us now see what the same texts would yield if analysed as two streams of speech. The texts can be first segmented into their constituent syllables so as to analyse their prosodic differences before they are considered as sequences of phonemes.

5.0.2.1 Syllabic Segmentation

A strange language may first strike one by its peculiar rhythm, the pattern of alternation between loud and quiet syllables, as is Hungarian, to a unilingual speaker of English. Or the first thing that strikes our ears may be the rise and fall of the voice, as Chinese or Swedish would strike those who do not know these languages. In each case, the first thing that attracts attention is not so much the quality of the individual sounds, but rather the tone and loudness of the syllables.

If the rhythm and intonation are not too different to what the hearer is used to, he may be able to spot some familiar words. For example, the word Canada in a Hungarian text may be twice as easy for an English speaker to spot as would the same word in a French text, although in most other respects French is more different from Hungarian than it is from English (See above).

Languages may sound more different in their rhythm and intonation than in their constituent phonemes. Because of this, homonyms may be all but unrecognizable when pronounced in another language. Compare, for example, F. développement / E. development

F. /devlop'mɑ̃/ / E. /də'veləpmənt/

Prosodic analysis includes features of rhythm and intonation; features of catenation may be handled more conveniently along with

the phonetic analysis. The unit of prosodic analysis is the syllable. Using the best fitting interfacing of syllables, based on the foregoing lexical matching pairs, in what respect, apart from its constituent phonemes, can one syllable differ from another?

Syllables can differ in loudness and length, on the one hand, and on height and direction of voice on the other. Recurrent changes of the first two features produce the audible differences in rhythm; variations in the second pair reveal differences in intonation. In order to be able to analyse and measure these differences, a simple notation for each feature and a corresponding numerical scale can be used (See 3.2.2 above).

Applying the notational scale, by way of example, to the above pair of sentences, one can measure the extent to which they differ prosodically.

Operations	Examples
1. Interface text equivalents and segment into syllables.	(A) Le patron a subi un mauvais coup. (X) The boss had a stroke of bad luck.
2. Interface syllables for best fit, e.g. stress opposite stress so that extra syllables are the weakest.	(A) Le | pa | tron | a | su | bi | un | mau | vais | coup (X) The | boss | had | a | stroke | of | bad | luck
3. Total blanks in bottom line. Total sums. This gives the difference in syllables. $L(\overline{A} \sim \overline{X})$	L | pa | tron | a | su | bi | un | mau | vais The | boss | had | a | stroke | of | bad coup 2 luck 4 $L(\overline{A} - \overline{X}) = 2$ $L(\overline{X} - \overline{A}) = 4$ $L(\overline{A} \sim \overline{X}) = 6$

4. Mark remaining syllables for sentence rhythm - force(f) and length(l), and intonation-height (h), and direction(d) of tone. (See Table 6.)

A ˌle ˌtron ˌa ˌmi ˌmau ˌcoup

X ˌthe ˌboss had ˌa ˌbad ˌluck

5. Number items on line a according to table of values. (See table.)

f (a)	1	1	1	1	1	2
l (a)	1	1	1	1	1	2
d (a)	0	0	0	0	0	0
h (a)	1	3	1	1	3	0

6. Repeat for line x.

f (x)	0	1	0	0	1	2
l (x)	0	1	0	0	1	2
d (x)	0	0	0	0	0	1
h (x)	0	2	0	0	3	3

7. Interface lines

1. f(a) - f(x)
2. l(a) - l(x)
3. d(a) - d(x)
4. h(a) - h(x)

f (a)	1	1	1	1	1	2	
f (x)	0	1	0	0	1	2	
f(a-x)	1	+0	+1	+1	+0	+0	= 3

f(a-x) = 3

l (a)	1	1	1	1	1	2	

8. Calculate difference.

l (x)	0	1	0	0	1	2	
l(a-x)	1	0	1	1	0	0	= 3

l(a-x) = 3

d (a)	0	0	0	0	0	0	
							= 1
d (x)	0	0	0	0	0	1	
d(a-x)	0	+0	+0	+0	+0	+1	= 1

d(a-x) = 1

h (a)	1	3	1	1	3	0	
h (x)	0	2	0	0	3	3	
h(a-x)	1	1	1	1	0	3	= 7

h(a-x) = 7

9. Compute total.

L ($\bar{a} \sim \bar{x}$) = 14

5.0.2.2 Phoneme Segmentation

Extending the analysis below the prosodic level, the phonetic transcriptions of two equivalent texts can be interfaced to compute the constituent and vectorial distances of their distinctive features. Continuing, therefore, with the above example, the texts in the phonetic notation may be processed as follows:

Operations	Examples
1. Transcribe equivalent sentences phonetically to include modifications of catenation.	lə patɚ̃ a sɥi œ̃ movɛ ku ðə lɒs had ə strouk ov bæd lʌk
2. Interface word and word-group equivalents.	lə patɚ̃ a sɥi œ̃ movɛ ku ðə lɒs had ə strouk ov bæd lʌk
3. In sequencial order, from left to right, interface phonemes for best fit within equivalent segments - vowels with vowels, consonants with consonants. Mark (/) if there is no opposite.	lə patɚ̃ /a sɥi œ̃ ///// // movɛ ku / ðə lɒs// ha //d/ ə strouk ov bæd/ lʌk 2 + 1 + 2 + 1 + 5 + 2 + 1 + 1 = 15
4. Total (/) differences.	L/a ~x/ = 15 (Sum of differences in number of phonemes per semantic equivalent).
5. Below each phoneme pair indicate inter-phoneme distance in vectorial values and add total. (See Tables 7 and 8.)	l ə p a t a b œ̃ m o v ɻ u o ə b p s ə d ə b æ d l ʌ 3+ 1 +1 +3 +1 +2 +3 +2 +1 +2 +3 +5 +3 = 30 E/a ~x/ = 30 (Sum of distances between matched phonemes.)

These interfacing texts may be considered either as static segments or as kinetic sequences. For there are, as we have seen, two types of measurement - static and kinetic, each applicable differently below the

prosodic level where distinctive sound features enter into the calculations. After the texts have thus been processed analytically, a few simple formulae can be applied for each of these types of measurement - static and kinetic. Let us begin with static measurement.

5.1 Static Measurement

What is here required is a formula for measuring the distance between two static interfacing texts as they appear to the eye. Since they will appear differently above the prosodic level than they will below, where the values of distinctive sound features appear, this demarkation will have to be taken into account. The first task is establishing an integrated formula for the measurement of distances in supra-prosodic segments.

5.1.1 Supra-prosodic Segments

One must begin, as above, by processing the text -- interfacing and numbering equivalent segments (N) from left to right and indicating their constituents (V). For example:

	1	2	3	4	5	
A.	Il	aime	beaucoup	la bière	de Munich./	N(a) = 5
X	He	likes	Munich	beer	very much.	N(x) = 5
	1	2	3	4	5	

N (a + x) = 10

Here the segmentation is for minimal equivalence. Although stratificational and transformation segmentation is also possible, these procedures may sometimes unnecessarily lengthen the analysis and computation.

The next step is to calculate the number of transpositions (T) necessary to make both structures congruent. This is done by matching the equivalent positions and adding their differences in natural numbers:

	1	2	3	4	5
A.	Il	aime	beaucoup	la bière	de Munich./
X.	He	likes	very much	beer	Munich.
	1	2	5	4	3
A.	1	2	3	4	5
X.	1	2	5	4	3

T(A-X) = 0 + 0 + 2 + 0 + 2 = 4

It is then possible to measure the difference between the constituent verbal elements (V) by counting the number of equivalent interfacing pairs with common one-to-one categories (c) and subtracting them from the total:

A.	Il	aime	beaucoup	la	bière	de	Munich./
X.	He	likes	very much		beer		Munich.
	2	2			2	2	

V(a) = 7
V(x) = 6
V(a+x) = 13
C(a+x) = 8

This difference in the make-up of the interfacing verbal constituents (V) can be expressed by the formula:

$$V(A\text{-}X) = V(a+x) - C(a+x).$$

Here, V = V(7+6) - C(2 2 2 2) = 13 - 8 = 5

Finally the difference between the units (U) comprising these constituents can be measured by subtracting the identical pairs (p) from the total number (n) of units, after interfacing the equivalents for maximal correspondence. This can be performed on texts in either orthographic or phonematic transcription. For example:

b i è r e = 5
| | n = 9
b e e r = 4

2 + 2 = p = 4

U = n - p = 9 - 4 = 5

/b jɛːʁ/ = 4
n = 8
/bɪəɹ/ = 4

2 p = 2

U = n - p = 8 - 2 = 6

M u n i c h = 6
| | | | | | n = 12
M u n i c h = 6

2+2+2+2+2+2 = p = 12

U = n - p = 12 - 12 = 0

/m y n i k/ = 5
| ( ) n = 11
/m ju n ɪ k/ = 6

2 + 2 + 2 = p = 6

U = n - p = 11 - 6 = 5

In this way distances of homographs and homophones can be distinguished. (The distance in the full orthographic texts comes out to 54 - 52 = 32.)

There are now three measures (U, V, and T) of the structures and constituents of the interfacing texts. These measures cannot be added because they represent essentially different things. Before they can be integrated they must first be reduced to the same scale.

This can be achieved by placing the minimum and the maximum differences in each case between 0 and 1. If the minimum in each case is fixed at 0, the maximum for U, V and T must then be determined. In the case of U, if all interfacing graphic forms are identical, then max U = n and $\frac{\max (U)}{n} = 1$.

In the case of V, if all equivalent constituents are compatible in category and level, then $\frac{\max(V)}{v} = 1$. Or, $\frac{\max (n - p)}{n} = 1$, and $\frac{\max(v - c)}{v} = 1$.

In the case of T, if equivalent segments occupy the same position then, min(T) = 0. But how is the maximum determined? If all interfacing segments are numbered as above, according to their position in the sequence, and their position numbers subtracted in positive numbers, a measure

of the structural difference between the texts is obtained. What is the limit or maximum to which such differences can extend? This maximum is obviously reached when the paired equivalents are interfaced in reverse order. For example:

```
                    1          2
F.  competence  administrative/

E.  administrative competence
                    2          1

                1   2   3
G.  Hier  bin  ich./

E.  I   am  here.
    3   2   1
```

Let us examine these relations and extend them so as to extract the pattern.

1 2 = 2 positions :

- 2 1

1 + 1 = 2 $\qquad \frac{2 \times 2}{2} = \frac{2^2}{2}$

1 2 3 4 = 4 positions :

- 4 3 2 1

3 + 1 + 1 + 3 = 8 $\qquad \frac{4 \times 4}{2} = \frac{4^2}{2}$

1 2 3 4 5 6 = 6 positions :

- 6 5 4 3 2 1

5 + 3 + 1 + 1 + 3 + 5 = 16 $\qquad \frac{6 \times 6}{2} = \frac{6^2}{2}$

N positions :

$$\frac{N \times N}{2} = \frac{N^2}{2}$$

Note, however, that in the above, N is always even.

Suppose that the total positions comes out to an odd number (3, 5, 7, ...). Examine the following:

```
  1
- 1
  0    =    0
```

1 position: $\frac{(1 \times 1) - 1}{2} = \frac{1^2 - 1}{2}$

```
  1   2   3
- 3   2   1
  2 + 0 + 2  =  4
```

3 positions: $\frac{(3 \times 3) - 1}{2} = \frac{3^2 - 1}{2}$

```
  1   2   3   4   5
- 5   4   3   2   1
  4 + 2 + 0 + 2 + 4  =  12
```

5 positions: $\frac{(5 \times 5) - 1}{2} = \frac{5^2 - 1}{2}$

N - 1 positions: $\frac{(N \times N) - 1}{2} = \frac{N^2 - 1}{2}$

The formula for the maximum number of structural transpositions, therefore, will depend on whether the total number of segments is odd or even. So that if N is even, $\max T(N) = \frac{N^2}{2}$; if N is odd, $\max T(N-1) = \frac{N^2 - 1}{2}$

Demonstration:

If $\max T(N) = \frac{N^2}{2}$ and $\max T(N-1) = \frac{N^2 - 1}{2}$, this should be valid for N, N + 1, N + 2, etc. Let us demonstrate the truth of this for N + 2.

$$\frac{N^2}{2} + 2(N + 2 - 1) = \frac{N^2 + 4N + 4}{2} = \frac{(N + 2)^2}{2}$$

$$\frac{N^2 - 1}{2} + 2(N + 2 - 1) = \frac{N^2 - 1 + 4N + 4}{2} = \frac{N + 1)(N + 3)}{2}$$

So that the ratio between the square of the total number of segments in two interfacing structures and twice the number of transpositions needed to make

then congruent ranges between 0 and 1. Therefore:

$$\max T(N) = \frac{N^2}{2} \quad \text{and} \quad \frac{2}{N^2} \max T(N) = 1$$

$$\max T(N-1) = \frac{N^2 - 1}{2} \quad \text{and} \quad \frac{2}{N^2 - 1} \max T(N-1) = 1$$

$$0 \leq \frac{2T}{N^2} \leq 1 \qquad \max\left(\frac{V}{v}\right) = 1$$

$$\max\left(\frac{U}{n}\right) = 1$$

The supra-prosodic distance (D), therefore, between two interfacing structures of Language A and Language X, is equal to twice the number of equalizing transpositions (2T) over the square of their combined equivalent segments (N^2) plus the difference between the constituent elements (V), plus the total graphic components minus the number of identical pairs, each set ranging in value between 0 and 1.

$$\tilde{D}\left(\frac{A}{X}\right) = \frac{2T}{N^2} + \frac{V}{v} + \frac{U}{n}$$

These integrated measures must therefore range between 0 and 3:

$$0 \leq \tilde{D}\left(\frac{A}{X}\right) \leq 3$$

But $\frac{U}{n} = \frac{n - p}{n}$. Therefore the integrated formula for static differences in texts analysed above the prosodic level must be:

$$0 \leq \tilde{D}\left(\frac{A}{X}\right) \leq 3 = \frac{2T}{N^2} + \frac{u - c}{u} + \frac{n - p}{n}$$

5.1.2 Infra-prosodic Segments

If interlingual distance differences below the prosodic level are to be included, we must measure features that are acoustic rather than visual and state them in a way that reflects their degree of similarities.

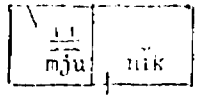
The acoustic features have to do with the prosodic variables of the syllable; the allophonic features, with differences in the pronunciation of phonemes in connected speech. This means that in the integrated formula, one term (the graphic/phonematic) will have to be replaced by two (the syllabic and the phonetic).

5.1.2.1 Syllabic Units

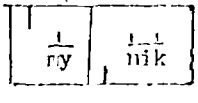
In adding a measure of prosodic distance to the formula, one must apply procedures of syllabic measurement before working out the formula for its integration.

Procedures of prosodic measurement, as already explained, include the interfacing of the syllables of equivalent speech segments for optimal correspondence and the marking of each syllable for rhythm (force and length) and intonation (height and direction). Each of these four sound components is measured on a scale of relative values for force (f), length (l), direction (d) and height (h) (See 4.2.0 above).

For example, take the homograph Munich as pronounced in comparable contexts in English and in French. First devide this stretch of speech into syllables and mark each for relative length, force, height and direction.

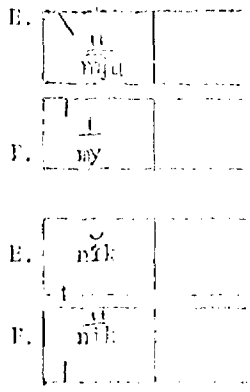
E. 

= h. 3 2 f 1 0
d 1 2 l 0 0

F. 

= h 3 1 f 2 2
d 0 1 l 0 1

Give the proper numerical values to the marked differences and treat each as a prosodic vector - subtracting one from the other in positive integers. For example:



E.		f	l	d	h	
		2	2	1	3	
F.		1	1	0	3	
		1	+1	+1	+0	= 3
E.		0	0	0	1	
F.		2	1	0	2	
	F =	2	+1	+0	+1	= 4/7

Then add the vectorial differences for each pair of syllables (3 + 4), to obtain the prosodic distance (7).

In brief, the prosodic distance between two equivalent stretches of speech is calculated by taking the sum (Σ) of the difference between each juxtaposed syllable (S) at each position (q) until all syllabic positions (Q) have been covered. That is:

$$\sum_{q=1}^{Q} S_q$$

In the above example, Q = 2 syllabic positions, $q_1 = 3$ and $q_2 = 4$.

$$\sum_{q=1}^{Q} = 3 + 4 \qquad \sum_{q=1}^{Q} S_q = 7$$

The next problem is to integrate this into the general formula. This can be done by expressing the syllabic distance as a percentage of the limit (L) of the maximum (max), which is 12.

Table 11

SIGNS AND SYMBOLS FOR INTEGRAL MEASUREMENT

Q = total of juxtaposed syllables (incl. empty apposition)	K = total of juxtaposed phonemes (incl. empty apposition)
q = a given syllabic position	k = a given phonemic position
S_q = syllabic difference at position q	P_K = phoneme difference at position k
L = upper differential prosodic limit	B = upper differential phoneme boundary
l = lower differential prosodic limit	b = lower differential phoneme boundary
U = units difference	u = number of units
V = constituent difference	v = number of constituents
C = common category	c = number of common constituents
N = number of segments	n = number of elements
T = transposition	p = number of matched pairs

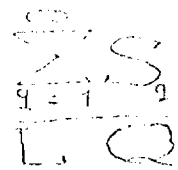
$$\max\left(\sum_{q=1}^{Q} S_q\right) = \sum_{q=1}^{Q} \max S_q = BQ$$

That is, the maximal possible difference in the prosodic features of any two stretches of speech (IQ) is equal to the total number of syllables (S) times the differential limit (L), or maximum difference between two syllables. This maximum must be made equal to 1 if the prosodic formula is to be integrated. So that,

$$\frac{1}{lQ} \max\left(\sum_{q=1}^{Q} S_q\right) = 1$$

This will have to be added to the infra-prosodic kinetic formula as a new term expressing the prosodic differences in the pronunciation of any two stretches of speech as being equal to the sum of the differences between all interfaced syllables $\left(\sum_{q=1}^{Q} S_q\right)$ over the differential prosodic limit (L) times the total

number of syllables (Q), that is,



Since psoition in a sentence is sometimes a function of intonation and rhythm, it may be useful to repeat the calculations by juxtaposing entire sentences as sequences of syllables.

5.1.3.2 Phoneme Units

Distance between the phonetic constituents of the syllables can be further measured. It is at this level that the effects of catenation, often analysed under the prosodic constituents of speech, can be taken into account.

For example, the French/English homograph observation could be transcribed, segmented and interfaced thus:

F.	ɔ	p	s	ɛ	ʁ	v	a	s	j	ɔ̃	/
E.	ɒ	b	z	ə		v	e	ʃ	ə	n	

Preliminary to any such measurement is an adequate transcription of the two interfacing texts. Under each interfacing pair, the vectorial differences are indicated; in the examples, these will be taken from the table of phonere distances (See Table 8). The next step is to take the sum of these distances as expressed in the formula:

$$\sum_{k=1}^{K} P_k$$

where K is the total number of positions; k, any given position, and P_k, the vectorial distance between any two phoneres interfaced at position k. Using a previous example for purposes of comparison:

F. bière / E. beer

k_1 k_2 k_3 k_4 = 4 = K

F. b j ɛ: ʁ

E. b I ə ɹ

0 + 2 + 2 +4 = 8

$$\sum_{K} P = \underset{k_1}{P} + \underset{k_2}{P} + \underset{k_3}{P} + \underset{k_4}{P} = \sum_{K=1}^{15} P_K = 8$$

To integrate this measure into the general formula, all possible results must be fitted between zero and one. This means that the calculations must be made within boundaries, a lower boundary (b) and an upper boundary (B). If two phonemes are phonetically identical, like the French and the English /m/, the vectorial distance between their prononciation is zero, which is the least, or lower boundary (b), and b = 0. The two phonemes of the interfaced phonetic codes which differ the most will have reached the upper boundary of possible difference between a phoneme in one language and a phoneme in the other language. In the case of French and English, it would, according to our table, be 17.5; but since in actual stretches of speech there can be context conditioned length differences, we must add another point to take such possible differences into account. So that B /F.~E./ = 18.5.

The maximum difference possible (BK) in the phonetic features of any two equivalent stretches of speech is equal to its length in the total number of positions (K) times the maximum difference between any two phonemes ($\max P_k$), each taken from one of the languages being compared. So that,

$$BK = \sum_{K=1}^{K} \max P_k .$$

In order to fit into our formula, this maximum must be made

equivalent to one. So that $\frac{1}{BK} \cdot \left(\sum_{k=1}^{K} P_k\right) \leq 1$

The phonetic difference between the pronunciation of any two stretches of speech is equal to the total number of interfacing phonemes $\left(\sum_{k=1}^{K}\right)$ (including nul opposites) times their interfaced differences (P_k) over their maximum vectorial distances (B), times the number of positions (K). That is, $\frac{\sum_{k=1}^{K} P_k}{BK}$. Applying this to the above example, where K=4, $P_k = 8$, B = 18.5, we get $\frac{\sum P_k}{BK} = \frac{8}{18.5 \times 4} = \frac{8}{74} = .108$

The general formula extended below the phoneme level would therefore require four terms:

1. $0 \leq \left(\frac{2T}{N^2} \text{ or } \frac{2T}{N^2 - 1}\right) \leq 1$

2. $0 \leq \left(\frac{u - c}{u}\right) \leq 1$

3. $\frac{1}{L} \leq \left(\frac{\sum_{q=1}^{Q} S_q}{LQ}\right) \leq 1$

4. $\frac{b}{B} \leq \left(\frac{\sum_{k=1}^{K} P_k}{BK}\right) \leq 1$

Since each of the above ranges between zero and one, the integrated formula must represent a range between zero and four. Here then is the integrated formula for static distance, extending below the phoneme level $\left(\frac{A}{\lambda}\right)$ between any two equivalent stretches of speech:

$$0 \leq \overline{D}\left(\frac{A}{\lambda}\right) \leq 4 = \frac{2\sum T}{N^2} + \frac{u - c}{u} + \frac{\sum S_q}{LQ} + \frac{\sum P_k}{BK}$$

$*\left(\frac{2T}{N^2 - 1}\right)$ if $c \neq 0$

5.2 Kinetic Measurement

In the integral measures of discourse distance, the interfacing texts have been considered as if they were two lines in space, segmented so that the equivalents can be matched and measured. In this static type of measurement, the time element of communication was not taken into consideration.

If the time element, or more exactly, the order of appearance of the constituent units, is taken into account, the segments take on the character of sequences. In finding the difference between two matching sequences of discourse, the interfacing constituents can be measured as being either the same or different, or as differing in degree.

In the first case, being above the prosodic (supra-prosodic) level, differences are discovered through the graphic representations of the texts - either orthographic or phonemetic. As we have seen, this graphic representation of speech may be limited to the sequence of phonemes; or it may be a very narrow transcription of the speech sounds.

5.2.1 Supra-Prosodic Sequences

The interfacing texts appear with elements arranged in a certain order;and since spatial order can be treated as analogous to temporal order, the same transposition formula already established for structural distance can be used, that is, $\frac{2T}{N^2}$

What must now be considered is not only what equivalent elements are identical, but where - or rather, when. That is, at what point in the time sequence does the identity appear? In other words, not only the number of positions but also their sequential order must be taken into account.

Suppose, for example, that one were to measure the kinetic

distance between the French word formules and the English (N. American) equivalent formulas, in phonematic (broadest) transcription:

F. f ɔ ʁ m y l (a)

E. f ɔ ɹ m ju l ə z (x)

(a ≤ x).

Examining the above interfaced transcriptions, we note the following:

1. The French form takes up six positions, whereas the English equivalent takes up nine.

2. Of the maximum of nine positions, six have interfacing equivalents; and three /j,ə,z/ have none. Let us label these the group K, so that $K = \{j,ə,z\}$.

3. Of the positions filled with a pair of interfacing phonemes, some are the same and some are different. Let us label this latter group L, so that $L = \begin{Bmatrix} ʁ & y \\ ɹ & u \end{Bmatrix}$.

4. Four of the phonemes as here represented are identical in both stretches of speech. This is the group $\{f,ɔ,m,l\}$, and it is obviously equal to zero, $\{f,ɔ,m,l\} = 0$. Now the differences need be measured.

5. We note that ʁ/ɹ occupy Position 3, and y/u occupy Position 6. So that $L\{3,6\}$ or $L = \{L_3, L_6\}$.

6. Each of these positions of the different interfacing units (a + x), i.e. (6 + 9) can be expressed as an average

$$\frac{a + x}{2} = \frac{6 + 9}{2} = \frac{15}{2}.$$

7. We can express the value of the common denominator by a factor equivalent to the number of interfacing pairs (2) in the sequence, i.e. $(1_3, 1_6)$ by considering them as exponents the values of which are reduced to 2 powers, e.g. (3-2, 6-2), giving us, $2^{3-2}, 2^{6-2}$.... or more generally, 2^{L-2}. Using this variable denominator for each unidentical pair thus gives values which decrease according to the distance of the unit from the first position, in this case, from the beginning of the word. So that R/J = $15/2^{3-2}$, and y/u = $15/2^{6-2}$, or generally $\frac{a + x}{2^{L-2}}$.

8. Looking at the first (Ø,x) set of words without opposites, $\{j,e,z\}$, or set K, occupying positions $\{k_5, k_8, k_9\}$, we need only vary the formula by reducing the power of the exponent by one, so that 2^{k-1} becomes the denominator. Therefore, $\{Ø/j\} = 2^{5-1}$, $\{Ø/e\} = 2^{8-1}$, $\{Ø/z\} = 2^{9-1}$, or generally $\frac{a + x}{2^{k-1}}$

9. The general formula for the kinetic distance between two equivalent interfacing sequences (W) is therefore:

$$W = \sum_{k \in K} \frac{a + x}{2^{k-1}} + \sum_{l \in L} \frac{a + x}{2^{l-2}}$$

So that $\frac{6 + 9}{2^{5-1}} + \frac{6 + 9}{2^{8-1}} + \frac{6 + 9}{2^{9-1}} + \frac{6 + 9}{2^{3-2}} + \frac{6 + 9}{2^{6-2}} =$

$W(\text{formyl}/\text{formulo}) = \frac{15}{16} + \frac{15}{128} + \frac{15}{256} + \frac{15}{2} + \frac{15}{16} = 9.106$

To get an idea of how the distances obtained through such measurement may differ from those obtained through static measurements and also between stretches of speech in orthographic and phonetic transcription, let us apply this formula to the above examples of F. bière and E. beer:

$l_1\ l_2\ l_3\ l_4\ l_5$

a = b i è r e = 5L

x = b e e r = 4L

$a + x = 5 + 4 = 9$

$K = \{e\} = 1^{k-5}$ $\qquad L = \{i/e,\ e/e\} = \{l_2,\ l_3\}$

$K = \{e/\emptyset\} = 2^{5-1}$ $\qquad L = \{?\}$

$$\sum_{k \in K} \frac{a+x}{2^{k-1}} = \frac{5+4}{2^{5-1}}, \quad \sum_{l \in L} \frac{a+x}{2^{l-2}} = \frac{5+4}{2^{2-2}} + \frac{5+4}{2^{3-2}}$$

$$\sum_{k \in K} \frac{a+x}{2^{k-1}} + \sum_{l \in L} \frac{a+x}{2^{l-2}} = \frac{9}{2^4} + \frac{9}{2^0} + \frac{9}{2^1} = \frac{9}{16} + \frac{9}{1} + \frac{9}{2} =$$

$$= .56 + 9.00 + .45 = 10.01$$

$$W\left(\frac{\text{bière}}{\text{beer}}\right) = 10.01$$

The same formula applied to a phonetic (narrow) transcription of these two words would show a greater difference, corresponding to the different pronunciations of the (r). It would show: (bjɛːʁ/ bɪəɹ)

$L_1\ L_2\ L_3\ L_4$

a = b j ɛː ʁ = 4

x = b ɪ ə ɹ = 4

$a + x = 8$

$K\ \{\emptyset\} = 0$ $\qquad L = \{l_2,\ l_3,\ l_4\} = 3$

$$\sum_{k \in K} \frac{a+x}{2^{k-1}} + \sum_{l \in L} \frac{a+x}{2^{l-2}} = 0 + \frac{4+4}{2^{2-2}} + \frac{4+4}{2^{3-2}} + \frac{4+4}{2^{4-2}} =$$

$$\frac{8}{2^0} + \frac{8}{2^1} + \frac{8}{2^2} = \frac{8}{1} + \frac{8}{2} + \frac{8}{4} = 8+4+2 = 14. \quad W\left(\frac{\text{b j ɛː ʁ}}{\text{b ɪ ə ɹ}}\right) = 14$$

To find the kinetic differences between two sequences of words, one must simply add the differences between each interfacing pair. The difference for a text of (I) pairs is the sum of the individual differences. That is,

$$\sum_{\lambda=1}^{I} W_{\lambda} = \sum_{\lambda=1}^{I} \sum_{\lambda \in K} \frac{\partial_{\lambda} + \chi_{\lambda}}{2^{k-1}} + \sum_{\lambda=1}^{I} \sum_{\lambda \in L} \frac{\partial_{\lambda} + \chi_{\lambda}}{2}$$

5.2.2 Infra-prosodic Sequences

Even though the transcription of the interfacing stretches of speech should be most narrow, the amount of detail needed to differentiate the sounds may not be sufficient. This is especially likely in the measurement of differences between closely related dialects and also in the study of bilingual interference. Below the graphic level of differentiation sounds can be broken down into distinctive features, each of which is scaled, as has been done for the infra-prosodic static measurements (See Tables 6, 7 and 8).

The difference here, however, is that the actual vectorial value of the distance between two interfacing sounds depends on its position in the sequence, which has been established as a minimal equivalence. As in the kinetic measurement of interfacing segments of graphic symbols (orthographic or phonetic) the values given to the differences are decremental - they get smaller as the sequence advances in time. Here,however, the differences must be recorded, not as same-or-different symbol, but as a degree of similarity or difference. This is obtained by subtracting distinctive features.

The treatment of these two types of measurement must therefore be somewhat different. Symbols that are the same, different, or without opposites need no longer be taken into account, since those representing sounds with identical features will have a distance of zero, those that are different will show a positive vectorial value, and those without opposites may be given either the value resulting from the sum of their distinctive

features, or half the maximum. Consequently, sequences of sounds will appear as sequences of numbers each representing the degree of difference from its opposite, according to its position. The position gives each number a decremental value changing the degrees of difference, which are then totaled and averaged.

In order successively to decrease the values of the vectorial distances, the powers of the positions (k) of the interfacing pairs (p), which increase with the length of the sequence (2^1, 2^2, 2^3, 2^4 ....2^k = 2,4,8,16 ...n) must be used correspondingly as devisors of these vectorial distances, which then become the dividends ($P^k/2^k$). Since the difference between two corresponding sounds is also a function of the length (K) of their equivalent sequences (i), this must also enter into the calculation: e.g. the o in pop/pop (2^3) cannot be given the same value as the o in population/population (2^{10}). This can be entered as a constant multiple of each distance and also as a divisor of the total. Let us see how this works out in practice.

Repeating one of the above examples, to facilitate comparison, measure the phonetic difference between the pronunciation of F. bière and E. beer, as follows:

1. Interface the phonemes for optimal sequential correspondence	k_1	k_2	k_3	k_4
	b	j	ɛ:	ʁ
	b / 0	f / 2	ə / 2	ɹ / 4
2. Note vectorial distances (P^k) See table for English/French.	0	2	2	4
3. Multiply each by 2 (for pair) raised to the same power as the length of the sequence (K_i), in this case, $K_i = 4$, i.e. $2^{ki} = 2^4$	k_1	k_2	k_3	k_4
	0.2^4	2.2^4	2.2^4	2.2^4
	0	32	32	64

4. Divide each of these products by 2 raised to the power of the position's ordinal ($_2$k). $\frac{0}{2^1}$ $\frac{32}{2^2}$ $\frac{32}{2^3}$ $\frac{64}{2^4}$

2 4 9 16

5. Total and divide by length of sequences as a power of the pair ($_2$ki = 4 = 16). 0 + 8 + 4 + 4 = 64

The kinetic vectorial distance between F./bjɛʁ/ and E. /bɪəd/ will then be 4. $\frac{64}{16} = 4$

This operation can be summarized in a formula for measuring the kinetic vectorial distance in the phonetics ($\underline{P}$) of two equivalent sequences (i):

$$\underline{P}_i = \sum_{k=1}^{k=n} \frac{P_k}{2^k}$$

Although one may not require such refined measurements in comparing extensive texts in two languages, a formula for such measurements may be needed for the exact calculation of the distance between two dialects or for the study of bilingual interference. By way of example, let us apply this formula to a transcription of an item in a tape recording of a bilingual Acadian recounting a story in English and using French place names. One of the French place names is Rivière-du-Loup, which in the speaker's own dialect of French is pronounced as [ʁ i v ɛ ʁ d y ˈl u] but in the English context, with stress-conditioned vowel conversion, as [ˈ ɹ i v ə ɹ d ə ɬ u] . Let us see what the kinetic difference is between these two stretches of speech by applying the above formula.

Λ (E_i) =	ɹ	i	v		ə	ɹ	d	ə	ɬ	u
X (F_j) =	r	ɪ	v	j	ɛ	r	d	y	l	u
	—	—	—	—	—	—	—	—	—	—
P_k =	6	1	0	12	1	6	.5	3	1	.5

$$P_i = \frac{6}{2^1} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{12}{2^4} + \frac{1}{2^5} + \frac{6}{2^6} + \frac{.5}{2^7} + \frac{3}{2^8} + \frac{1}{2^9} + \frac{.5}{2^{10}}$$

$$= \frac{}{2^{10}}$$

$$\underline{P}_i = \frac{3072 + 256 + 0 + 768 + 32 + 96 + 4 + 12 + .5}{1024}$$

$$\underline{P}_i = \frac{4242.5}{1024}$$

$$\underline{P}_i = 4.4$$

This gives a measure applicable to each interfacing equivalent sequence of speech. To apply it to a text (e.g. a sentence or series of sentences) containing a number (J) of consecutive stretches, interfaced after transposition for minimal equivalence, one needs an over-all formula to add the distances in the corresponding sequences of which the texts are composed, that is, $\sum_{i=1}^{J}$.

So that, $$\sum_{i=1}^{I} \underline{P}_i = \sum_{i=1}^{J} \sum_{k=1}^{K_i} \frac{\underline{D}_k}{2^k}$$

5.2.3 Integration

The above kinetic formulae measure the distance between two equivalent minimal stretches of discourse as each of their constituents (phonemes or letters) appears in time. But how about the order of such stretches? Since these can be treated as being analogous to the static segments, the same formula may be applied $(\frac{2T}{N^2})$. In order to add its value, which ranges between 0 and 1, the kinetic formulae would have to be adapted to it, so that they too would have the same range. How can this be done? Let us first integrate the phonetic measures.

Since the upper and lower boundaries (b/B) do not change, they can be multiplied for each stretch by 1 minus 1 over 2 to the power of the length, as indicated by the last ordinal in the sequence of positions (K_i), that is $\left(1 - \frac{1}{2^{K_i}}\right)$ for the total number of stretches of speech $\sum_{i=1}^{I}$ and $\sum_{i=1}^{I}\left(1 - \frac{1}{2^{K_i}}\right)$.

So that,

$$\frac{\max\left(\sum_{\lambda=1}^{I}\sum_{\lambda=1}^{K_\lambda}\frac{P_\lambda}{2^\lambda}\right)}{B\sum_{\lambda=1}^{I}\left(1-\frac{1}{2^{K_\lambda}}\right)} = 1$$

The integrated kinetic formula for the vectorial distance between two oral texts $\left(\frac{A}{X}\right)$ therefore, is:

$$\overline{D}\left(\frac{A}{X}\right) = \frac{2T^{\times}}{N^2} + \frac{\sum_{\lambda=1}^{I}P_\lambda}{B\sum_{\lambda=1}^{I}\left(1-\frac{1}{2^{K_\lambda}}\right)}$$

$\times \frac{2T}{N^2-1}$ if N is odd.

Finally, let us work out the formula for integrating kinetic distances between two graphic texts. If the texts are considered as graphic sequences, the lower boundary is zero, since $\min(W_\lambda) = \sum_{\lambda=1}^{I}(n-n)W_\lambda = 0$, if $K_\lambda = L_\lambda = \{\emptyset\}_\lambda$

For the upper boundary, all interfacing items must be different $L_\lambda \{1,\ldots 2_{\lambda-1}\}$

So that,

$$\max(W_\lambda) = \sum_{\lambda=1}^{X_\lambda}\frac{X_\lambda}{2^{\lambda-1}} = X_\lambda\sum_{\lambda=1}^{X_\lambda}\frac{1}{2^{\lambda-1}} = X_\lambda\left[8-2^{3-X_\lambda}\right]$$

$$\max\left(\sum_{\lambda=1}^{I}W_\lambda\right) = \sum_{\lambda=1}^{I}(\max W_\lambda) = \sum_{\lambda=1}^{I}X_\lambda\left[8-2^{3-X_\lambda}\right]$$

$$\max\left(\sum_{\lambda=1}^{I}W_\lambda\right) \Big/ \sum_{\lambda=1}^{I}X_\lambda\left[8-2^{3-X_\lambda}\right] = 1$$

The integrated kinetic formula for distances between two graphic texts would therefore be:

$$\overline{D}\left(\frac{A}{X}\right) = \frac{2T^{\times}}{N^2} + \sum_{\lambda=1}^{I}W_\lambda \Big/ \sum_{\lambda=1}^{I}X_\lambda\left[8-2^{3-X_\lambda}\right]$$

This last formula is, of course, less important than the one for the measurement of kinetic distance in speech, since the time element affects the written word less. The position of an item in a sequence, however, may be the deciding factor in interlingual identification and interference, since speech is essentially a chain of associations. It is for the study of such questions as these that this type of interlingual distance may be worth measuring.

CONCLUSION

The measurement of interlingual distance is both possible and feasible; and it can be computed in different ways. The difference between the codes of the two languages can be measured by one technique and the differences in samples of discourse by another. The samples may be measured as static entities covering a certain space, or as dynamic or kinetic sequences unrolling in time. The distance between two languages may be measured as the sum of their differences or as the amount of work necessary to convert one language into another. Each can be measured either by taking all the characteristics in which two languages can differ and counting the number of differences in each, or by integrating the immediately observable differences into a single formula of measurement.

Measurement in the differences in the semantic organization of languages can be kept separate from measurement of differences in their formal features; or equivalence of meaning can be considered as a constant of interfacing and equivalent texts. Distance between languages can also be measured indirectly by comparing each with a standard form, or both with a third language.

The type of measurement used will depend on the purposes and reasons

for wanting to measure the distance between two languages. The purposes may be practical or theoretical. The practical purposes are very much like our needs for any sort of measurement in our daily lives - measurements of time, temperature, weight, size and speed, which one continually encounters. The usefulness of such measures depends on whether they are sufficient for the purpose. An outside house thermometer, for example, is an instrument sufficient to let a person know whether or not to dress warmly on the way to work in the morning. But it may not be sufficient to let that same person know whether or not one of the children has to stay away from school that day because of a fever. Whether the weather or the fever is measured in centigrade or farenheit scale may be pointless to argue; all that may be needed is to know which is which.

In the practical study of language related problems, wide-mesh investigations will call for different types of measures than will fine-mesh laboratory types of research. And studies of language acquisition may require types of measures different from those suited to studies of bilingual and bi-dialectical interference.

The effects produced by contact of two highly similar languages may be quite different from those produced by the contact of two highly different languages. In order to be able to measure the relative distribution of each language in the behaviour of a bilingual, we must first be able to distinguish between his languages. For purpose of analysis and measurement, it is not sufficient to know which languages these are; one must also have information on the particular dialect or dialects used and the extent of difference between them.

On the theoretical plane, interlingual measurement can contribute to the solution of general problems in linguistics and psychology. Since many of

these are problems of categorization - of same or different and more or less - their solution may lie in substituting scales for the categories now in use, that is, by making the differences measurable.

One of these problems, as we have seen, is the existence of universal categories of language - grammatical, lexical and phonetic. Have all human beings common characteristics of mind, body and environment which create universal language categories in their codes, their discourse and their systems of communication? Beyond the self-evident and axiomatic of logic, how can one prove that a language category is universal? In the process of verification it may become evident that certain language categories are more general than others, and the degree of generality may apply to more important categories than their either/or universality. In sum, linguistics may not be able to solve the problems of language universals until it can express itself in the most universal language of them all - the language of number.