The Economics of Artificial Languages: An Exploration in Cost Minimization

Jonathan Pool

Fachtagung Angewandte Soziolinguistik,
Paderborn, Germany, 19-20 June 1980

Utilika Foundation
Seattle, Washington
http://utilika.org

FEoLL

THE ECONOMICS OF ARTIFICIAL LANGUAGES:

AN EXPLORATION IN COST MINIMIZATION

Jonathan Pool

Institut für Kybernetische Pädagogik
Forschungs- und Entwicklungszentrum für
objektivierte Lehr- und Lernverfahren
Postfach 1567
4790 Paderborn 1, Federal Republic of Germany
(until 14 August, 1980)

and

Department of Political Science, DO-30
University of Washington
Seattle, Washington 98195, USA

Prepared for delivery at Fachtagung Angewandte Soziolinguistik,
Universität Paderborn, 19-20 June, 1980

The Economics of Artificial Languages: An Exploration in Cost Minimization

Jonathan Pool

1. Artificial Codes, Artificial Languages, and Natural Languages

The term "artificial language" is used in several senses, of which there are two major ones in scientific contexts. First, it can mean what I shall call an "artificial code": a set of explicit rules for communication which have come into existence by being prescribed. Such codes have been devised for communication among humans, between humans and animals, between humans and machines, between animals and machines, among animals, and among machines. The secret codes invented by pickpockets or baseball players, and the public codes embodied in the colors of armed forces uniforms, are examples of artificial codes for human-human communication. Chimpanzees have been taught artificial codes (Rumbaugh, 1977). And hundreds of programming languages have been invented for human-machine communication (Feldman, 1979).

As used below, "artificial language" will have a second sense, differing from "artificial code". A tentative definition is: a partially explicit set of rules for communication among humans, which has evolved from an artificial code, which is usable both orally and in writing, and which allows itself to be modified to communicate any ideas that can be communicated with any natural language. In the history of the world approximately one thousand artificial codes are known to have been devised with the expressed hope that they would evolve into artificial languages (Duličenko, n.d.). A few of them have actually done so to a noticeable degree; these include Volapük, Esperanto, and Ido. Today Esperanto, in fact, ranks approximately 65th among the world's languages in its rate of book publication ("Laste aperis", 1979; United Nations, 1979: 943-945).

The distinction between artificial codes and artificial languages follows Cherry's (1966: 8) distinction between sign systems and languages. Artificial languages, as defined above, resemble one of seven types of languages (standard, classical, artificial, vernacular, dialect, creole, and pidgin) that have been identified by Stewart (1968) on the basis of the presence or absence of four characteristics: standardization, autonomy, historicity, and vitality. In particular, artificial languages, according to Stewart, have standardization and autonomy but lack historicity and vitality. In other words, they are (in part) formally codified and their evolution is not dominated by any other language, but they are not perceived as resulting from a long period of development and they are not primarily transmitted over time as native languages. However, if an artificial language were to obtain historicity and vitality after some time, it would no longer be artificial in Stewart's typology, while it would still be artificial according to my definition. Except for this difference, the term "natural language" used above is intended to correspond to any of Stewart's other six types.

2. The Economic Importance of Artificial Languages

Both artificial codes and artificial languages have important economic aspects. Among the former, it is well known that programming languages differ in the efficiency with which they can obtain certain responses from certain computers. The optimal choice of programming languages requires taking into account the costs of language learning, programming, program modification, computer-language translation, computer execution, program storage, data storage, and conversion of programs across time- and machine-specific versions of the language or across languages (cf. Pratt, 1975: 6-10).

It is less known how artificial languages, such as Esperanto and Volapük, differ in their economically relevant characteristics, but a number of studies have found major differences in such characteristics between an artificial language and a natural one (e.g. French). In spite of this, there is little theory about the economics of human languages in general, and almost none about the economics of artificial languages. This is surprising when one considers that far more resources are certainly spent on the learning and use of human languages (exempting native languages) than on the learning and use of programming languages.

The paucity of glottoeconomic theory has the result that we know little about how to analyze the decisions that political authorities and ordinary individuals make about language: decisions about which languages to learn, which languages to teach, which languages to officialize, which language reforms to carry out, etc. Decisions like these have economic consequences for multilingual countries (India, USSR, Canada, etc.), for countries that use writing systems of questionable efficiency (Chinese, Japanese, English, Arabic, etc.), and for organizations whose members are highly diverse in their linguistic repertoires (United Nations, European Communities, Organization of African Unity, Union of International Associations, World Council of Churches, etc.). Such decisions impact not only total expenditures and total benefits, but also the way costs and benefits are distributed. The absolute and relative welfare of entire countries, social classes, professions, and ethnic groups is affected, and the effects may be large (see, e.g., Fishman, Ferguson, and Das Gupta, 1968; Frank, 1971; Rubin and Jernudd, 1971; United Nations, 1977).

3. Alleged Economic Advantages and Disadvantages of Artificial Languages

It may be possible for a person, organization, set of persons, or set of organizations to obtain economic benefits of various kinds by choosing an artificial language rather than a natural language when making certain language decisions. Let us survey the types of benefits that have been alleged to exist.

a. The learning of an artificial language is less costly than the learning of a natural language. A number of studies in various countries have evaluated this claim. Although precision in the comparison of learnabilities is difficult to obtain, although most of the studies until now have been weak in the degree of experimental control achieved, and although all studies conducted so far of which I am aware have used one and the same artificial language (Esperanto), still the results have been reasonably consistent in supporting the claim. Reported and estimated ratios of learning times for the achievement of similar levels of (active and passive) competence in Esperanto vs. natural languages have ranged from 1:2.7 up to 1:15 (Frank, 1977; Janton, 1973: 118; Kalckhoff, 1978; Markarian, 1964: 6; Rakuša, 1970: 38-39). When the research literature is combined with statements made by speakers of Esperanto whose native languages are non-Indo-European (e.g. El Popola Ĉinio, 1980), it appears that the ratio of learning times for Esperanto vs. natural Indo-European languages does not differ substantially between native speakers of Indo-European languages and native speakers of other languages.

Because of the dominant share that Indo-European languages have in the second-language market, little is known about the relative learnability of Esperanto vs. non-Indo-European languages as second languages. Wells (1978) has argued that Esperanto is an Indo-European language only in its lexicon, and that its phonology, morphology, syntax, and semantics contain attributes resembling several other language families. Whatever lexical bias toward the Romance and Germanic languages exists in Esperanto or any other artificial language may be substantially smaller than the corresponding bias of a natural language because of the high productivity of the affixes and grammatical morphemes and the wide international currency of the lexical morphemes of the artificial language. Thus undocumented claims that, for example, Esperanto cannot be learned any faster than English by a native speaker of Fijian (e.g. Farb, 1977: ch. 16) are clearly premature. Further research, in which cross-cultural and cross-lingual comparisons are made among learners, in which prestige differences between the target languages (presumably unfavorable to the artificial one) are reduced, in which experimenter biases (presumably favorable to the artificial language) are reduced or balanced, and in which different modes and degrees of competence are examined, will help resolve this question.

b. Translation and interpretation from or into an artificial language is less costly than from or into a natural language. There are three major reasons for expecting that this claim may be true. First, if an artificial language is less costly to learn than a natural language, and if the cost of translation and interpretation reflects the cost of training required to create the manpower that performs this service, then the cost of translating or interpreting

<u>from</u> an artificial language should be lower than the cost of translating or interpreting from a natural language, since in general translators and interpreters work from a second language into their native one. Second, considerable (sometimes nearly 100%) waste of translating and interpreting manpower has been reported in international organizations as a result of the unpredictability of the demand for translating and interpreting services, the rapidity with which the value of their products declines with time, the high cost of transportation among and temporary lodging at the sites of meetings, the fine linguistic specialization of existing translators and interpreters, and the failure of simultaneous interpretation (especially when conducted through an intermediate language) to convey more than a fraction (sometimes estimated at 50%) of the meaning into the target language (Piron and Tonkin, 1979; United Nations, 1977). If an artificial language can be learned well enough to be used as a target language as well as a source language, then the choice of an artificial language as target language does not constrain the choice of translators or interpreters to those having a particular native language, and the waste due to the "lumpiness" of this service, unpredictability, poor quality, etc. can be reduced. Third, it can be expected that an increasing proportion of the cost of translation in the future will be attributable to computer-initiated or computer-assisted translation. The difficulty and hence cost of most aspects of the computer processing of Esperanto has been reported to be less than that of computer processing of natural languages (Maas, 1975). This is because the artificial language, being syntactically, morphologically, and semantically more regular and less ambiguous, permits analysis with smaller programs, which therefore cost less to

develop, store, and execute, and which result in target-language texts that require less human correction to render them acceptable.

c. The learning of a natural language costs less when it begins with the learning of an artificial language than when it begins without such a preparatory phase. According to this claim, the learning of an artificial language causes learners to increase their aptitude and desire for natural-language learning. Under some conditions (e.g. 500 or more hours of study of a natural language) the investment in the artificial-language preparatory phase speeds the subsequent rate of natural-language learning enough to more than repay the time invested in the preparatory phase (Frank, 1978; Frank, Geisler, and Meder, 1979; Geisler, 1980; Markarian, 1964). Hence two languages--a natural and an artificial--can be learned at a lower total cost than a natural language alone. Total cost must, of course, also include the costs of teacher preparation, teaching materials, etc.; but Lobin (1978) estimates these at only about $1 per pupil per year more for a two-language program than for a one-language program because of the extremely short time (1 week) required to retrain foreign language teachers to teach an additional course in an artificial language. Reasons for the impact of an artificial language on the subsequent learning of a natural language may include (a) an increased understanding of grammatical concepts resulting from the learning of a language which has a high degree of one-to-one correspondence between grammatical morphemes and grammatical concepts, (b) an increased motivation for second-language learning resulting from an initial experience of success with an easily learned language, (c) an increased proclivity to learn the most useful lexical items as a result of learning a language a high proportion of whose lexicon is of this kind (Chaves, 1979), or (d) the increased aptitude that the learning of any second language confers. If the last reason is the predominating mechanism, then any other language would do as well as an artificial one, unless the determining variable is the amount of second-language competence rather than the amount of second-language study. Research comparing an artificial and a natural language (i.e. a different natural language from the target one) as linguistic propaedeutics has apparently not yet been conducted.

d. The choice of an artificial rather than a natural language as the medium of communication in situations where there is only one such medium and in which competence in the medium is a prerequisite for engaging in a certain kind of production results in a more efficient allocation of manpower among productive activities. The argument is that there are many persons with high aptitudes in certain fields (e.g. natural sciences and engineering) but low aptitudes in second-language learning, since these two kinds of aptitudes are not strongly correlated. When the language of trans-national communication in a field of production is a natural one, a larger number of persons who are fitted by aptitude for that field are excluded from it (or from prominence in it) by failing to achieve competence in that language than is the case when the language is artificial and hence not as difficult to learn. I have not discovered any statistics on the rate of disqualification for professional study on the basis of failure to satisfy foreign or national language requirements, but this appears to be a major phenomenon in countries where a non-native language is the major medium of higher education (e.g. India) or is used as a higher education selection criterion even though not used as a medium of instruction at that level (e.g.

Turkey). Noss (1967) has dealt with these problems in the countries of Southeast Asia. On the other hand, selection of an artificial language for professional communication which has previously been served by a natural language entails, as does any change of language, a period of bilingual transition, in which it is even more costly than before for the linguistic needs of the profession to be met (by language learning or translation of literature). How costly this transition period would be would depend on the rate of obsolescence of the literature and of the manpower in the field.

e. The choice of an artificial language as one medium of communication in situations where those competent in any of the permitted media of communication enjoy substantial advantages creates a political situation in which a smaller number of languages can be chosen, a smaller amount of translation and interpretation is required, a smaller total expenditure is required, a more equal sharing of the total expenditure is obtained, a larger total benefit from communication is obtained, and a more equal sharing of that benefit is achieved. These advantages of an artificial language are claimed to result from what is called its "neutrality". The argument is generally applied to the choice of official languages in international organizations some of whose members are native speakers of the current or prospective official languages. In that case, any one official language would advantage the member that speaks it natively, and the prospect of such an advantage induces other members to oppose the officialization of that language without their own language also being officialized. Thus the inexorable increase in the number of official languages that has been observed in the United Nations system (United Nations, 1977) emerges. The officialization of an artificial language would not lead to such a process, since no member would be substantially and obviously advantaged thereby. This argument, however, would not apply under certain conditions: (1) where the artificial language has obtained large numbers of native speakers and they are unevenly distributed among the members of the organization; (2) where at least one candidate for officialization among the natural languages has very few or no native speakers among the members (e.g. a multilingual state with a history of a foreign official language); (3) where the members are willing to accept the principle of non-linguistic (e.g. monetary) compensation for linguistic disadvantages.

Five alleged economic advantages of artificial over natural languages, and qualifications of them, have been summarized above. There are others in the economic realm, such as the lower cost of programming a speech synthesizer to convert a standard written text into intelligible speech (Sherwood, 1979). There have also been, of course, assertions of psychologically, socially, and politically relevant differences, which will not be considered here. In spite of the primitive state of current knowledge, it appears that artificial languages have been and therefore can be devised in such a way that they can be learned, or translated from, at substantially less than the cost incurred for learning or translating from a natural language. A difference of such magnitude deserves efforts at empirical verification and explanation and, in the meantime, some analysis of how language decisions should rationally be made if two kinds of languages really exist with such major differences.

4. The Problem of Language Choice

Let us now look at some implications of different learning costs between artificial and natural languages. What follows will be only a small part of the analysis that the subject requires. We shall focus on the problem of choosing languages as media of communication between those who do not yet share a language. We shall ask which choices of language minimize the total cost incurred by a set of people. We shall ignore the possibility of indirect communication, such as through translators and interpreters. We shall also refrain from examining the political problem of inducing people to cooperate in bringing about the cost-minimizing solution. In fact, we shall not even take the benefits of communication into account when weighing the costs of achieving it. Thus we shall be following only a few of the many implications of the differences between artificial and natural languages.

To represent these implications I shall use a system of formal notation called "APL", while also explicating most of the APL expressions in plain English. This will allow the formulas that are derived below to be handled directly by a computer when they are applied to specific problems.

First, let us describe the situation. At any given time, there is a relevant world, and it consists of a number of members. We might think of this world as a community, and of the members as people who live in it. But we could also apply such an analysis to an international organization whose members are delegations, to an international federation whose members are national associations, etc.

We can describe the relevant world with numeric variables. A variable describing a single person normally has a value consisting of just one number, e.g. that person's age or authoritarianism score. For the relevant world, however, a variable may need a value composed of many numbers, such as one for each member or one for each possible pair of members.

A most important variable for describing the relevant world will be one that represents the language repertoires of the members of the world. Let us call it REP. If for the sake of simplicity we assume, initially, that a given member either knows or does not know a given language, i.e. that there are no intermediate levels of language competence, then REP can be a logical matrix, i.e. a rectangular array of 1's and 0's, with one row for each member and one column for each language in the relevant world. For example, if REP is

$$\begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{matrix}$$

then members 1 and 4 know only language 2, member 2 knows only language 1, and member 3 is bilingual, knowing languages 1 and 2. There are 3 monolinguals, 1 bilingual, and no trilinguals. Two members know language 1, 3 members know language 2, and no member **knows** language 3. There is one language with no speakers, there is no language with one speaker, there is one language with 2 speakers, there is one language with 3 speakers, and there is no language with 4 speakers. All these facts about the world are implicit in the above matrix.

Once REP, the language repertoire variable, is known, we can derive from it another logical variable, which represents which pairs of members share at least one language and can therefore communicate directly with each other. Let us call this variable DIR. DIR is defined as

$$\text{DIR} \leftrightarrow \text{REP} \vee . \wedge ⍉ \text{REP}$$

which means that DIR has a 1 at row i and column j if and only if there is at least one column of REP in which rows i and j both have a 1. Otherwise DIR consists of 0's. The example of REP above would yield this DIR:

1	0	1	1
0	1	1	0
1	1	1	1
1	0	1	1

Among the many sentences that could be uttered on the basis of this variable, we see that member 3 can communicate directly with all other members, and member 2 can communicate directly with only one other.

Now we are ready to introduce the cost of language learning. For any member and any language, it would cost a certain amount for that member to learn that language. We can represent this set of learning costs as a variable LCO, and it will have the same shape as REP. An example of LCO that is consistent with REP above is

10	0	1
0	20	4
0	0	2
16	0	2

Naturally, the cost is represented as 0 whenever the member already knows the language. This example of LCO shows what the situation might be like if languages 1 and 2 were natural and language 3 were artificial. The cost of learning language 3, in this example, ranges from 20% down to 10% of the cost of learning language 1 or 2, depending on who the learner is and which natural language the artificial one is being compared with. In general, we notice substantial cost differences among the members in this example, such as the fact that it would cost member 4 twice as much to learn language 3 and 1.6 times as much to learn language 1 as it would cost member 1. What do these differences mean? One interpretation is that these are truly comparable absolute costs, arrived at by some objective definition, such as a weighted sum of hours and money. Let us call this assumption "absolute comparability". Another interpretation is that the absolute costs are not comparable among members, since they are based on the subjective utilities to each member of time, money, effort, etc. Under the second interpretation, we can still say that the relative cost of learning language 1 vs. language 3 is higher for member 1 than for member 4, but we could not say that the absolute cost of learning language 1 is higher for member 4 than for member 1. LCO could be transformed by multiplying any row by a constant, e.g.

50	0	5
0	5	1
0	0	3
8	0	1

and it would not contain any different information, according to the second interpretation. Let us call this interpretation "relative comparability".

When we assume absolute comparability, it is possible to add the costs incurred by several members, deriving a total cost; it is also possible to compute what proportion of the total cost is borne by each member. This allows us to determine which of several alternatives has the lowest total cost, which alternative distributes its cost most equally, etc.

We can now see how the costs of alternative means of establishing direct communication can be compared, assuming for the moment absolute comparability. The total cost of establishing the possibility of direct communication between any two members, i and j, can be represented by a variable called i TOT j. Its value is a set of L numbers, where L is the number of languages in the relevant world (i.e. the number of columns in REP). i TOT j (or j TOT i, which is identical) can be derived simply from LCO, by summing rows i and j. In APL notation,

i TOT j ↔ +⌿LCO[i j;]

In the example given above, members 2 and 4 are currently unable to communicate directly. When we use the second version of LCO above, 2 TOT 4 has the value 8 5 2, which means that the total cost of enabling them to communicate in language 1, 2, or 3 would be 8, 5, or 2, respectively. In this case, the cost would be minimized if they both learned the artificial language. But now consider members 1 and 2. 1 TOT 2 is 50 5 6, which means that, even though the artificial language is at least 5 times less costly for members 1 and 2 to learn than the natural language which they do not yet know, it is still best (i.e. total-cost minimizing) for them to establish communication by member 2 learning (natural) language 2, because member 1 has much higher learning costs than member 2. Finally, 2 TOT 3 has a value 0 5 4, which tells us that language 1 is the cost-minimizing language for them to communicate in. The reason, of course, is that they both already know language 1.

We can define i MTL j as the language through which a communication possibility between members i and j can be established at minimum total cost. Thus, if 2 TOT 4 has the value 8 5 2, then 2 MTL 4 has the value 3. The formal definition of i MTL j is

i MTL j — ω X=⌊/X← i TOT j

where ωV is synonymous with V/ιρV when V is a logical one-dimensional array (vector).

In the general case, then, where absolute comparability is assumed and where the alternatives for establishing communication possibilities are considered for only 2 members at a time, the total-cost-minimizing alternative is given by i MTL j, which is identical to j MTL i. When only 2 members are considered at a time, the situation is particularly simple. Each alternative for enabling those members to communicate consists of a single language. The separate costs to the 2 members are represented by their respective rows of LCO, and the total costs are given by adding these two rows together, column by column.

The situation becomes more complicated when we ask about the costs of the various alternatives for bringing a set of W members into a state of mutual direct communicability, where W > 2. The number of alternatives rises rapidly with W, and the costs to any one member are no longer representable as a row of LCO. One practical way to evaluate the alternatives is to follow three steps: (1) determine what the alternatives are; (2) determine for each member what the costs of all the alternatives are; (3) add these costs together, alternative by alternative, to obtain the total cost for each alternative, and find the smallest of these. As a by-product, this leaves us with separate cost comparisons for the individual members, which will be useful for other kinds of analysis.

Suppose we list the members among whom we desire to enable communication and call this list V. An example of V is

1 2 4

Step 1 above can be accomplished by defining a variable ALT V, which represents a list of all the alternative ways that the pairs of members in V who cannot already communicate directly could be enabled to communicate. ALT V is defined as

```
ALT V ↔ 1+(N ρ L)⊤ ¯1+ ιS←(L←1 ↓ ρREP)*N←1 ↓ρ P←((,⍉C),[.5],
C←( ρB) ρ D)[; ω ,(D∘.<D← ι W← ρ V) × ∼B←DIR[V;V]]
```

which, in the example we have been using, would have the value

1	1	1	2	2	2	3	3	3
1	2	3	1	2	3	1	2	3

ALT V in this example has 2 rows because there are 2 pairs of members in V who cannot yet communicate: members 1 and 2, and members 2 and 4. The first row represents the first of these pairs, and the second row the second. The numbers represent the languages in which these pairs could communicate if one or both of them learned those languages. Thus the 6th column, for example, represents the alternative in which members 1 and 2 learn to communicate in language 2, and members 2 and 4 learn to communicate in language 3. Inspection of REP reveals that this alternative would require member 1 to learn nothing, member 2 to learn languages 2 and 3, and member 4 to learn language 3.

Once ALT V has been computed, it is possible to carry out step 2 by finding what each alternative would cost each affected member. The variable that lists these separate costs can be called SEP V, and one way to define it is with a function (i.e. program) that computes the costs one member at a time:

```
    Z ← SEP V
[1] Z ← (W,S) ρ C ← ⍬
[2] Z[C;] ← (∨⌿ALT[⍵∨⌿P=C;] °.= ⍳ L)+.× LCO[V[C ← C+1 ];]
[3] → (C < W)/2
```

Applying this function to the second cost variable LCO above, we would obtain this value for SEP 1 2 4:

50	50	50	0	0	0	5	5	5
0	5	1	5	5	6	1	6	1
8	0	1	8	0	1	8	0	1

The 3 rows of SEP 1 2 4 represent the costs of the 9 alternatives for the 3 members in V who are not already able to communicate with everyone else in V (which in this case happens to be all the members in V), members 1, 2, and 4. For instance, the 6th alternative, mentioned above, would cost member 1 nothing, member 2 6, and member 3 1.

From here it is a simple matter to perform step 3, obtaining the total costs. Let us call the total cost variable TOC V. It is defined as

```
TOC V ↔ +/SEP V
```

and in the above example has the value

```
58 55 52 13 5 7 14 11 7
```

If we let the alternative which can enable all the members in V to communicate at the lowest possible cost be the value of a variable MTA V, it can be defined as

```
MTA V ↔ P,[1](⍉ALT V)[ω X=⌊/X ← TOC V;]
```

In our example, MTA 1 2 4 would have the value

```
1 2
2 4
2 2
```

where in each column the first two numbers designate a pair of members and the third indicates what language that pair should use. The value shown above means that the use of language 2 by members 1 and 2 and also by members 2 and 4 would be the least expensive way of bringing about total communicability among members 1, 2, and 4.

This method of finding the least expensive alternative for enabling direct communication among all pairs in a particular set of members can be applied to large numbers of people, although the computing formulas presented above will have to be revised when the computer's memory capacity is reached. This can easily happen, as becomes clear when one realizes that in a group of 6 members with 3 languages there can be more than 14 million alternatives. One plausible way to work with small numbers is to treat all persons having identical language repertoires as a single member. The number of persons in a member can then be reflected in the language-learning-cost variable LCO, if absolute comparability is assumed.

5. Language Choice in the Light of Artificial-Natural Differences

So far the analysis has been general with respect to whether the L languages that exist in the relevant world are natural or artificial. But whereas natural languages (except for pidgins and creoles) are generally viewed as being easy or difficult to learn largely as a result of their distance from one another, (some) artificial languages are asserted to be intrinsically very easy to learn, in addition to being easier for some to learn than for others because of linguistic distance. Let us then make the assumption that the relevant world has some number of natural languages and one artificial language. Let us further assume that the cost to member i of learning natural language p is a function of four things: (1) the number of people in i, (2) the distance of p from whichever language already known by i is the nearest to p, (3) the number of languages i already knows, and (4) as discussed in section 3 above, whether or not i already knows the artificial language. The cost to member i of learning the artificial language is a function of four things: (1) the number of people in i, (2) the artificial language's intrinsic learnability, (3) its distance from the nearest language that i already knows, and (4) the number of languages i already knows. If we assume that the innate capacity to learn languages is identically distributed in all language-repertoire groups and that each member in our analysis is a group of substantial size sharing a language repertoire, then we do not need to assume that members differ in their innate language-learning ability.

Let us here apply these assumptions to the case in which the relevant world contains 2 natural languages. All the possible language repertoires can be represented by REP as follows, where the third column represents the artificial language:

0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

Rows consisting of 0 0 0 and 0 0 1 are not included, because no normal person is alingual and no normal person knows only an artificial language yet. To determine the costs of language learning, we can let

N_i = the number of persons in member i

D_{ij} = the distance between languages i and j

E3 = the intrinsic learnability (ease) of language 3, the artificial language

A3 = the cost advantage that results from knowing an artificial language when learning a natural language

K2 = the cost advantage that results from knowing two languages compared with one language when learning any other language

Since REP is symmetrical with respect to languages 1 and 2, let us arbitrarily assume that language 2 is at least as close to language 3, linguistically, as is language 1 to language 3.

These assumptions allow us to determine a set of general formulas for LCO, the variable of language-learning costs:

N1 × D12	0	N1 × D23 × E3
N2×(D12⌊D13) ×A3×K2	0	0
0	N3 × D12	N3 × D13 × E3
0	N4×(D12⌊D23) ×A3×K2	0
0	0	N5 × D23 × E3 × K2
0	0	0

(A ⌊ B means the lesser of A and B.)

Before any further language learning takes place, the possibilities of direct communication are represented by DIR, whose value is

```
1 1 0 0 1 1
1 1 0 1 1 1
0 0 1 1 1 1
0 1 1 1 1 1
1 1 1 1 1 1
1 1 1 1 1 1
```

DIR makes it clear that there are only 3 pairs of members that cannot communicate directly: members 1 and 3, 1 and 4, and 2 and 3. Consequently, there are 27 alternatives that could enable all members in the relevant world to communicate directly, as given by ALT:

```
1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3
1 1 1 2 2 2 3 3 3 1 1 1 2 2 2 3 3 3 1 1 1 2 2 2 3 3 3
1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
```

We can also see that the cost to member 5 of learning language 3 is not relevant, since member 5 can already communicate with everyone else. Thus 6 learning costs remain in the analysis.

Another simplification emerges when we inspect the 27 alternatives in ALT. All but 6 of them are superfluous, because they either are synonymous or oversolve the problem by providing more media of communication (at a higher cost) than necessary. Let us redefine ALT to be the remaining set of 6 distinct and minimally sufficient alternatives, i.e.

1	1	1	2	2	3
1	1	1	2	3	3
1	2	3	2	2	3

In the eyes of the artificial-language planner, the crucial question is whether the conditions are auspicious for the choice of the artificial language as a solution (alternatives 3 and 5) or *the* solution (alternative 6) to the language problem. The members who do not yet know the artificial language (members 1 and 3) would not be willing to learn it voluntarily on the basis of cost alone, since doing so would clearly be more costly to them than sitting back and waiting for the other members to learn their natural languages (i.e. member 1 waiting for members 3 and 4 to learn language 2, and member 3 waiting for members 1 and 2 to learn language 1). But the planner might still show that total cost would be minimized by the use of the artificial language.

The 6 alternatives in the condensed ALT have the following total costs, as would be reflected in TOC 1 2 3 4 (here, as elsewhere, all operations in APL notation are executed strictly from right to left):

(1) $(N1 \times D12)+N2 \times (D12 \lfloor D13) \times A3 \times K2$
(2) $D12 \times N2+N3$
(3) $(N1 \times D12)+N3 \times D13 \times E3$
(4) $(N3 \times D12)+N4 \times (D12 \lfloor D23) \times A3 \times K2$
(5) $(N3 \times D12)+N1 \times D23 \times E3$
(6) $E3 \times (N1 \times D23)+N3 \times D13$

Given these costs, what kind of artificial language would succeed in making alternative 3, 5, or 6 the least costly?

It is reasonable to assume that, of all the cost components, two are within the power of the creator of an artificial language to manipulate (to some degree): the intrinsic learnability of the language (E3) and its relative distances from the relevant natural languages (D13 and D23). We can then ask what values these variables must have if an artificial-language alternative is to be optimal.

First, let us assume that the language has a particular location in the linguistic attribute space, so D13 and D23 are fixed. How does the optimal solution vary with E3? We can begin by comparing the three alternatives that make use of the artificial language. The all-artificial solution (alternative 6) will be better than either mixed solution (alternatives 3 and 5) if

$$(E3 \times (N1 \times D23) + N3 \times D13) < (N3 \times D12) + N1 \times D23 \times E3 \text{ and}$$
$$(E3 \times (N1 \times D23) + N3 \times D13) < (N1 \times D12) + N3 \times D13 \times E3$$

and hence if

$$E3 < D12 + D13 \text{ and}$$
$$E3 < D12 + D23$$

So let us here assume that both these inequalities are satisfied, i.e. that the all-artificial alternative is the only artificial one that may be optimal. How easy does the artificial language have to be to make that alternative the least costly? Its cost must simply be less than the costs of alternatives 1, 2, and 4, i.e. the following inequalities must be satisfied:

$$E3 < ((N1 \times D12) + N2 \times (D12 \lfloor D13) \times A3 \times K2) + (N1 \times D23) + N3 \times D13$$
$$E3 < (D12 \times N2 + N3) + (N1 \times D23) + N3 \times D13$$
$$E3 < ((N3 \times D12) + N4 \times (D12 \lfloor D23) \times A3 \times K2) + (N1 \times D23) + N3 \times D13$$

Now let us make the opposite assumption: that the learnability of the artificial language, E3, is given. How does the optimal solution vary with D13 and D23, the distances of the artificial language from the two natural languages? First, we can rephrase what we have already found about the relative costliness of the three solutions involving the artificial language. The all-artificial alternative, number 6, will be less costly than alternatives 3 and 5 if

$$D13 < D12+E3 \text{ and}$$
$$D23 < D12+E3$$

In other words, the all-artificial solution will be better than either partially artificial solution as long as the artificial language is close enough to both natural languages. The required proximity to the natural languages varies in direct proportion to the difficulty of the artificial language and in inverse proportion to the distance of the natural languages from each other. (A reasonable limit on proximity to the natural languages is also furnished by our spatial analogue: the path between the natural languages through the artificial one should be at least as long as the shortest path between them, i.e. $D13+D23 \geqslant D12$.)

If again we assume that the artificial language meets the conditions that allow us to disregard the partially artificial alternatives, we can determine what distances between artificial and natural languages would make the all-artificial solution optimal. The five inequalities that must be satisfied for this to be the case are:

$$((D13 \times N3)+D23 \times N1) < (D12 \times N1+N2 \times A3 \times K2)+E3$$
$$((D13 \times N3-N2 \times A3 \times K2)+D23 \times N1) < (N1 \times D12)+E3$$
$$((D13 \times N3)+D23 \times N1) < (D12 \times N2+N3)+E3$$
$$((D13 \times N3)+D23 \times N1) < (D12 \times N3+N4 \times A3 \times K2)+E3$$
$$((D13 \times N3)+D23 \times N1-N4 \times A3 \times K2) < (N3 \times D12)+E3$$

The coefficients of D13 and D23 in the second and fifth inequalities, respectively, can be negative under some conditions. When this is the case, the cost advantage of the artificial-language solution is actually reduced as a result of proximity between the artificial language and the natural ones. The reason is that this proximity facilitates the learning of the other natural language by members who already know the artificial one.

An example that applies these formulas will conclude this presentation. Suppose that the relevant world consists of people who speak Eurian, people who speak Afrasian, and people who speak both, and that a small proportion of each of these groups also speak Artifese. The figures are:

N1 = 10	D12 = 1
N2 = 1	D13 = .8
N3 = 40	D23 = .6
N4 = 2	E3 = .2
N5 = 8	A3 = .9
N6 = 2	K2 = .8

REP, DIR, and ALT have been presented above. The matrix of language-learning costs, LCO, for this example, is

10	0	1.2
.576	0	0
0	40	6.4
0	.864	0
0	0	.768
0	0	0

The total costs of the six alternatives, TOC 1 2 3 4, are

10.576 50 16.4 40.864 41.2 7.6

Thus the least costly alternative is for the monolingual Eurian speakers and Afrasian speakers to learn Artifese. A close second is for the Eurian speakers, whether or not they know Artifese, to learn Afrasian. This solution would surpass the artificial-language solution if any of several parameters were to change to a sufficient degree; for example, if the intrinsic learnability of the artificial language were to fall so that

$$E3 > .278316$$

If E3 were .3 instead of .2, for example, then the total costs would be

10.576 50 19.6 40.864 41.8 11.4

and the positions of the two least costly alternatives would be reversed, giving the advantage to Afrasian.

6. Conclusion

It would be foolish to suppose that, upon being informed that the artificial language's intrinsic learning cost had been remeasured as 30% rather than 20% of that of a natural language, the speakers of Eurian would throw away their Artifese textbooks and begin studying Afrasian. This mental experiment illustrates the need for a comprehensive analysis that includes not only total costs but also individual member costs, and not only costs but also benefits. The analysis begun here can be extended to model the common situation in which a central authority has some power to make collective language decisions (e.g. on which languages shall be taught in public schools), but individual members also have much decision-making power and exercise this power in light of the costs and benefits to themselves, plus their anticipations of the decisions that other members are going to make, rather than in order to minimize the total cost to society. If successful, the models that are thus developed should be able to explain and predict linguistic equilibria and change. They should also be able to assist those who make language decisions in better processing and evaluating complex information.

AUTHOR NOTE

I am grateful for research support from the Program in Comparative Studies in Ethnicity and Nationality, the Graduate School, and the Graduate School Research Fund of the University of Washington, and from the Institut für Kybernetische Pädagogik of the Forschungs- und Entwicklungszentrum für objektivierte Lehr- und Lernverfahren, for facilitating this work, as well as for helpful references to previous work from Prof. Dr. Helmar Frank.

REFERENCES

Chaves, Sylla M. (1979). Überlegungen zur Lernerleichterung im Fremdsprachunterricht durch Vorausstellung der Internacia Lingvo. Grundlagenstudien aus Kybernetik und Geisteswissenschaft, 20, 122-130.

Cherry, Colin (1966). On Human Communication, 2nd edn. Cambridge, Mass.: M.I.T. Press.

Duliĉenko, Aleksandr (n.d.). Bibliografija mezhdunarodnykh jazykov. Unpub. ms.

El Popola Ĉinio (1980).

Farb, Peter (1977). Word Play. New York: Bantam.

Feldman, Jerome A. (1979). Programming Languages. Scientific American, 241:6 (Dec.), 80-93.

Fishman, Joshua A.; Ferguson, Charles A.; and Das Gupta, Jyotirindra; eds. (1968). Language Problems of Developing Nations. New York: Wiley.

Frank, Helmar (1971). Die Sprachbarriere zwischen den Wissenschaftlern. Umschau in Wissenschaft und Technik, Heft 7, 236-238.

Frank, Helmar (1977). Die Meßmöglichkeit der relativen Schwierigkeit von Fremdsprachen. In Kybernetik und Bildung III, ed. Günter

Lobin and W. D. Ekkehard Bink (Paderborn: Schöningh/Hannover: Schroedel), 14-25.

Frank, Helmar (1978). Grundlagen und sprachpädagogische Anwendung einer informationstheoretischen Transferanalyse. Grundlagenstudien aus Kybernetik und Geisteswissenschaft, 19, 75-88.

Frank, Helmar; Geisler, Evelyn; and Meder, Brigitte S. (1979). Nachweise des strukturbedingten Transfers aus dem Sprachorientierungsunterricht. Grundlagenstudien aus Kybernetik und Geisteswissenschaft, 20, 14-28.

Geisler, Evelyn (1980). Mezurado de la lernplifaciligo de la angla pro ILo. Eŭropa Dokumentaro, 25, 4.

Janton, Pierre (1973). L'Espéranto. Paris: Presses universitaires de France.

Kalckhoff, Gerhard (1978). Eine vergleichende Darstellung der Erlernbarkeit von Sprachen. Grundlagenstudien aus Kybernetik und Geisteswissenschaft, 19, 55-60.

Laste aperis (1979). Esperanto, 72, passim.

Lobin, Günter (1978). Bildungsökonomische Analyse zu verschiedenen Modellen des Früh-Fremdsprachunterrichts. Grundlagenstudien aus Kybernetik und Geisteswissenschaft, 19, 126-132.

Maas, Heinz Dieter (1975). Zum Stand des automatischen Textworterkennung. Grundlagenstudien aus Kybernetik und Geisteswissehschaft, 16, 29-40.

Markarian, Raif (1964). The Educational Value of Esperanto Teaching in the Schools. London: Universal Esperanto Association Research and Documentation Centre.

Noss, Richard (1967). Higher Education and Development in Southeast Asia: Language Policy. Paris: UNESCO.

Piron, Claude, and Tonkin, Humphrey (1979). Translation in International

Organizations. Rotterdam: Universala Esperanto-Asocio.

Pratt, Terrence W. (1975). Programming Languages: Design and Implementation. Englewood Cliffs, N.J.: Prentice-Hall.

Rakuša, Rudolf (1970). Metodiko de la Esperanto-Instruado, 2nd edn. Ljubljana: Mladinska Knjiga.

Rubin, Joan, and Jernudd, Björn H., eds. (1971). Can Language Be Planned? Honolulu: University Press of Hawaii.

Rumbaugh, Duane M., ed. (1977). Language Learning by a Chimpanzee. New York: Academic Press.

Sherwood, Bruce Arne (1979). Schnelle textgesteuerte Sprachsynthese-Algorithmen für Esperanto, Spanisch, Italienisch, Russisch und Englisch. Grundlagenstudien aus Kybernetik und Geisteswissenschaft, 20, 109-121.

Stewart, William S. (1968). A Sociological Typology for Describing National Multilingualism. In Readings in the Sociology of Language, ed. Joshua A. Fishman (The Hague: Mouton), 531-545.

United Nations (1977). General Assembly. 32nd Session. Report on the Implications of Additional Languages in the United Nations System. Document A/32/237. 11 Oct.

United Nations (1979). Statistical Yearbook, 1978. New York: United Nations.

Wells, John (1978). Lingvistikaj aspektoj de Esperanto. Rotterdam: Universala Esperanto-Asocio.

KNAPPTEXT

Wirtschaftliche Auswertung von Plansprachen: Zum Problem der günstigsten Lösung

Jonathan Pool

Plansprachen haben eine kaum untersuchte aber immer noch erhebliche wirtschaftliche Bedeutung. Gegenüber natürlichen (ethnischen) Sprachen wird von Plansprachen behauptet (und teilweise empirisch nachgewiesen), daß sie

(1) mit weniger Aufwand gelernt werden können,
(2) schneller und leichter zu übersetzen sind,
(3) das spätere Lernen der natürlichen Sprachen beschleunigen,
(4) die Wirkung des Arbeitsmarkts reibungsloser machen würden und
(5) es erlauben würden, offizielle Mehrsprachigkeit durch eine billigere gemeinsame Zweitsprache zu ersetzen.

Als Beispiel für die als Thema gewählte Problematik wird folgende Frage behandelt: wie kann man am günstigsten, d.h. mit den niedrigsten Gesamtkosten, Sprachbarrieren überwinden? Die sprachlichen Gegebenheiten werden durch eine Reihe von Matrizen dargestellt: eine Sprachbeherrschungsmatrix, eine Sprachgemeinsamkeitsmatrix und eine Sprachlernkostenmatrix. Auf dieser Basis werden (DV-freundliche) Formeln abgeleitet, um die möglichen Lösungen und deren Kosten (getrennt für jedes Mitglied des Systems sowie die Gesamtkosten) zu ermitteln.

Anschließend wird diese Methode auf den Fall eines Systems angewandt, das zwei natürliche Sprachen und eine Plansprache enthält. Angenommen und als Modellparameter eingegeben werden bestimmte Verhältnisse zwischen--einerseits--Sprachabständen, der Anzahl beherrschter Sprachen, sowie der Art (natürlich oder geplant) der beherrschten und der zu lernenden Sprachen, und--andererseits--den Kosten des Sprachlernens. Dieses Verfahren ergibt Ungleichungen, die bestimmen, wie lernbar die Plansprache sein muß, um die günstigste Lösung darzustellen, und wie der sprachliche Abstand zwischen der Plansprache und den natürlichen Sprachen die Kostenverhältnisse beeinflußt. Es wird gezeigt, daß der Kostenvorteil einer Plansprache bei kleinerem Abstand zu den natürlichen Sprachen u.U. nicht verbessert sondern verringert werden kann.

RESUMO

Ekonomio de Planlingvoj: Kelkaj Kalkuloj pri Kosta Malplejigo

Jonathan Pool

Planlingvoj, simile al sed distingite de (ekz. komputorprogramaj) kodoj, estas ekonomie gravaj. Laŭ pluraj studoj kaj/aŭ argumentoj, planlingvoj kostas malpli ol naturaj lingvoj:

(1) Planlingvoj pli rapide lerneblas;
(2) ili pli facile tradukeblas;
(3) lerni planlingvon pliigas onian kapablon poste lerni naturajn lingvojn;
(4) profesia komunikado per planlingvo pliigas la efikecon de la labormerkato;
(5) oficialigi planlingvon politike ebligas forigi superfluajn oficialajn lingvojn kaj ties kostojn.

Unu aspekto de la ekonomia malsameco inter planaj kaj naturaj lingvoj analiziĝas ĉi tie: la kostoj de la du specoj. En la mond-modela matrico ekzistas nombro da anoj (vicoj) kaj nombro da lingvoj (kolonoj); iu ano aŭ scias (1) aŭ ne scias (0) iun lingvon. De tiu matrico produkteblas komunikmatrico, kiu montras kiuj havas komunan lingvon kun kiuj. Lernkosta matrico montras kiom kostus al iu ano lerni iun lingvon. Se oni supozas la kostojn al diversaj anoj kompareblaj kaj do sumeblaj, oni povas kalkuli la tutajn kostojn de la diversaj metodoj (t.e. kombinoj de lingvolernaj decidoj) por komunikivigi iun ajn aron da anoj kaj do trovi la solvon kiu havas la malplejan koston. La necesaj formuloj, rekte uzeblaj per komputoro, prezentiĝas.

Fine aplikiĝas tiuj formuloj al modelo de 2 naturaj kaj unu plana lingvoj. La lernkostoj estas funcio de interlingvaj distancoj, la jam sciata lingvonombro, ĉu oni jam scias planlingvon kaj ĉu la lernendaĵo planlingvas. La rezultoj montras kiom facile lernebla la planlingvo devas esti por plej malmultkosti. Alia formulo montras kiomaj distancoj povas ekzisti inter la plana kaj naturaj lingvoj sen difekti ĝian plejmalmultkostecon. Rezultas, ke proksimeco inter planlingvo kaj naturaj lingvoj povus eĉ malhelpi al planlingva solvo esti la malplejkosta.