

Interlingua and MT, a Discussion

by Jared Darlington *, Research Laboratory of Electronics,
Massachusetts Institute of Technology

This paper discusses a proposal by Alexander Gode that Interlingua be used as an intermediate language for mechanical translation. The word-by-word translations proposed by Gode from Interlingua into English are not always easily understandable or editable, because of the presence in Interlingua of idioms, reflexive verbs, multiple meanings for particles and other words, and non-English word-order. Some revisions in Interlingua are suggested which would make it more useful for mechanical translation.

In the December, 1955, issue of *MT*, Dr. Alexander Gode claims that "... a base text in Interlingua is convertible by mechanical means into an editable translation in a target language belonging to the group of languages which are summarized in Interlingua".* This "group of languages" includes primarily English, French, Italian, Spanish and Portuguese, and secondarily or derivatively Latin, Russian and German (*vide* the *Interlingua-English Dictionary*, N. Y., Storm, 1951). In the *MT* article, "mechanical" (i.e. word-by-word, or rote) translations are made from a source text in Interlingua into English, French and German. Though the results of these translations are not correct or idiomatic English, French or German, Gode believes them good enough to permit an editor (presumably monolingual) easily to transform them into correct, idiomatic language. There is no doubt that the sample translations which Gode presents are easily redactable, but in one sense they are oversimplified in that only one target-language equivalent is listed for each Interlingual word. In a strictly rote translation, many possibilities must be listed for words like 'de,' 'per,' and 'que,' and in translating these words respectively as 'of,' 'by,' and 'which,' Gode does not explain why he chooses these in preference to other possibilities like 'from,' 'through,' and 'that.' A program for the automatic englishing of Interlingua must either list all the English equivalents of each Interlingual word it encounters, or it must be able to decide, on the basis of contextual hints, which translation is most appropriate. That it will not suffice to proceed in an entirely word-by-word fashion, listing all entries for each word, may be readily seen by considering the following rote translation of an Interlingual sentence:

AT/TO + LESS + THAN/THAT/WHAT/WHICH/WHO/
WHOM + THE + GREAT + POWERS + WANT/WANTS/
WISH/WISHES + TO SAY/TO TELL + IT + THAN/THAT/
WHAT/ WHICH/ WHO/ WHOM + THEM/ THEY + SAY/ SAYS/

* This work was supported in part by the National Science Foundation and in part by the U.S. Army Signal Corps, the Air Force Office of Scientific Research, and the Office of Naval Research.

* Gode, Alexander, "Signal System in Interlingua," *Mechanical Translation*, Vol. 2, No. 3, p. 90 (1955).

TELL/TELLS + ABOUT/ABOVE/CONCERNING/ON/ON TOP/
ON TOP OF/OVER/UPON + THE + ENDING/TO END/FIN-
ISHING/TO FINISH + BELONGING TO THE/BY MEANS OF
THE/FROM THE/MADE OF THE/OF THE/SINCE THE/WITH
THE + PROOFS/TESTS/TRIALS + NUCLEAR + , + MANY/
MUCH + COUNTRIES/LANDS + MORE/PLUS + LITTLE/
SMALL + HERSELF/ HIMSELF/ ITSELF/ ONESELF/ THEM-
SELVES + WILL + FRIGHTEN.

The Interlingual sentence that gives rise to this farrago is:

A menos que le grande potentias vole dicer lo que illes dice super le finir del provas nuclear, multe paisas plus parve se espaventara.

In plain English, this means:

Unless the great powers mean what they say about the ending of nuclear tests, many smaller countries will be frightened.

The almost total unintelligibility of the sample rote translation is due to the many idiosyncrasies of Interlingua that are present in the original sentence. Among these are: the idiomatic nature of 'a menos que' ('unless'), 'vole dicer' ('mean'), and 'lo que' (relative pronoun 'that which' or 'what'); the reflexive nature of the verb 'se espaventar' ('to become frightened'); the multiple uses of the prepositions 'a,' 'de,' and 'super;' the substantive nature of 'finir,' requiring the English gerundial 'ending' (or 'finishing'); and the nonexistence of personal and numerical forms for the Interlingual verbs. Less serious are the departure from English word-order in 'provas nuclear' and 'paisas plus parve,' and the multiple entries for 'provas,' 'multe,' 'paisas,' 'plus,' and 'parve.'

The possibility of finding or constructing troublesome Interlingual sentences of this sort entails of course that this language as it stands is not a satisfactory source-language for rote translation into English. In this paper we propose to examine the idiosyncratic features of Interlingua in a little more detail, and to try to see what can be done about them. Since Interlingua is to some extent an artificially constructed lan-

guage, there is always the possibility of modifying it so as to eliminate various difficulties that crop up, an alternative that most definitely is not open in dealing with natural languages. For Interlingua too there is a limit, albeit vaguely defined, to the amount of permissible tampering, namely, Interlingua must not be made so like one of the contributing natural languages that it becomes too unlike one or more of the others. That is, its character as the “least common denominator” or “intersection” of the important western European languages must in some sense be preserved. In making Interlingua more “logical” so as to facilitate mechanical translation out of it, we must not make it so “unnatural” that it cannot easily be read by people with a “standard average European” (in Whorf’s sense) linguistic background.

Turning our attention next to the idioms* of Interlingua, we may divide them roughly into six categories: †

1. Idioms which can be literally translated into English with no loss of original meaning (strictly speaking, these interlinguicisms are not idiomatic with respect to English), such as:

- a. abundar in = to abound in
- b. cader malade = to fall ill
- c. esser curte de = to be short of
- d. esser tote aures = to be all ears
- e. in le calor de = in the heat of
- f. in le ultime analyse = in the final analysis
- g. justo nunc = just now
- h. sin dubita = without doubt

2. Idioms which can be literally translated into English, making only minor changes, with no loss of original meaning, such as:

- a. calefaction central = central heating
- b. critar al lupo = to cry wolf
- c. de tote lateres = on all sides
- d. esser de accordo = to be in accord
- e. fortia brute = brute force
- f. jocar de parolas = to play on words
- g. lassar multo a desirar = to leave much to be desired
- h. loco commun = commonplace

3. Idioms which, if literally translated, make sense but the wrong sense, such as:

* The following is a representative selection, rather than a complete listing, of Interlingual idioms. The sources for them, as well as for the other features of Interlingua discussed, are the Interlingua publications of Dr. A. Gode and associates, especially the *Interlingua-English Dictionary*, the *Interlingua* grammar (both N.Y., Storm, 1951), and *Novas de Interlingua*.

† We are not presupposing any particular definition of ‘idiom.’ An excellent discussion of the problem of defining this term may be found in Dr. Bar-Hillel’s paper, “Idioms,” in W. N. Locke and A. D. Booth, *Machine Translation of Languages*, N. Y., John Wiley & Sons, Inc., 1955. Bar-Hillel rightly points out that a distinction must be drawn between monolingual and bilingual idioms, and that no expression is ever idiomatic in an absolute sense, its idiomaticity being relative *inter alia* to a grammar and to a dictionary.

- a. de bon corde = gladly, willingly, *not* of good heart
- b. foras de se = beside oneself, *not* outside of oneself
- c. guardar le lecto = to stay in bed, *not* to guard the bed
- d. voler dicer = to mean, *not* to want to say

4. Idioms which, if literally translated making only minor changes, make sense but the wrong sense, such as:

- a. a fortia de = by means of, *not necessarily* by force of
- b. manducar le parolas = to mumble, *not* to eat one’s words
- c. societate anonyme = limited company, *not* anonymous society

5. Idioms which can be literally translated, but which have some English meanings that are not correct, such as:

- a. deponer un summa super un cosa = to put a sum on something (i.e., to bet, *not* to make a down payment)
- b. esser in balancia = to be in balance (i.e., to be undecided, *not* to be steady)
- c. prender le aer = to take the air (i.e., to get some fresh air, *not* to speak over the radio, or to leave)

6. Idioms whose literal translations are nonsensical, such as:

- a. a fin que = in order that
- b. a menos que = unless
- c. de hic a un hora = an hour from now
- d. experto contabile = accountant
- e. haber loco = to take place
- f. il conveni de facer le = it is advisable to do it
- g. il se tracta de = it is a matter of
- h. le unes le alteres = each other

Various proposals have been made for handling idioms in mechanical translation, and they often involve using a special idiom dictionary (*vide* Bar-Hillel, *op. cit.*). But there are two main difficulties in the use of an idiom dictionary, namely, (1) the existence of discontinuous idioms, as in ‘The Count di Luna got, or so he thought, *his own back*,’ and (2) the fact that certain expressions are sometimes idiomatic, sometimes not, as in ‘*In truth*, he has lost his faith.’ Mechanical means of handling discontinuous idioms and sometime-idioms are not in principle impossible to devise, but it would certainly be simpler if the source-language contained some further indications of the presence of idioms. As far as Interlingua is concerned, we may simply stipulate that no idioms are to be discontinuous, and further that all the words making up an idiom are to be connected either by hyphens (as in the English ‘to-day’ and ‘week-end’) or by outright compounding (as in ‘today’ and ‘weekend’). Thus, in Interlingua, we will get hyphenated expressions such as ‘a-menos-que’

('unless'), 'il-se-tracta-de' ('it is a matter of'), and 'a-fin-que' ('in order that'), or compound words, such as 'amenosque,' 'ilsetractade,' and 'afinque.' The ease of reading should be the crucial factor in deciding whether these idioms should occur as hyphenated or as compounded. For a rote translation routine, all that matters is that they not consist of words separated by spaces. The Interlingua dictionary will have to include these hyphenated or compounded idioms. Thus, the original writer of an Interlingua article or summary will do a certain amount of automatic "pre-editing" of his own work.

Turning our attention next to the reflexive verbs of Interlingua, we note that several of these do admit of a literal translation into English. For example:

- a. assecurar se que = to assure oneself that
- b. blandir se = to flatter oneself
- c. contentar se con = to content oneself with

Others yield wrong meanings under literal translations:

- a. affliger se = to grieve, *not* to afflict oneself
- b. batter se = to fight, *not* to beat oneself
- c. espaventar se = to become frightened, *not* to frighten oneself
- d. facer se tarde = to be late, *not* to make oneself late (being late is not always one's own fault)
- e. occupar se de = to be interested in, *not* to occupy oneself of

Still others yield no sensible literal translations:

- a. addormir se = to fall asleep
- b. affollar se = to get angry
- c. amicar se = to make friends
- d. debatter se = to argue
- e. obstinar se a = to persist in
- f. sentir se ben = to feel well

It would obviously simplify matters if the reflexive pronoun 'se' were always connected to the verb, by an apostrophe or by a hyphen. Thus, instead of 'ille se batte' we would have 'ille s'batte,' or 'ille se-batte.' Then, the correct translation 'he fights' would always result, and there would be no chance of ever getting the malapropos 'he beats himself.'

As for the prepositions and other grammatical words of Interlingua the main trouble is that one word is frequently used to signify several essentially different relations or concepts. The preposition 'de' is perhaps the worst offender, but is by no means the only one, some other culprits being:

- per = by, by means of, during, per, through, throughout
- perque = because, why
- post = after, afterwards, back, backwards, behind
- super = about, above, concerning, on, on top, on top of, over, upon.

The problems caused by the multiple entries for these and other grammatical words are compounded by the

fact that one word may perform several different syntactical feats, e.g., 'post' may be either an adverb or a preposition; 'perque' may be either an adverb or a conjunction; 'omne' may be either an adjective or a pronoun; 'ancora' may be either an adverb or an interjection; 'alique' may be either an adverb or a pronoun 'que' may be either a conjunction, an interrogative pronoun, or a relative pronoun; 'bastante' may be either an adjective or an adverb; and so it goes. There is also in many cases a confusion between a spatial and a temporal sense, as in 'ante,' which as a preposition can mean either 'in front of (in space) or 'before' (in time) and which as an adverb can mean either 'ahead' (in space) or 'earlier' (in time). In a case like this, one might conceivably argue that there is no important difference among these four senses, and that Interlingua is quite right to summarize them all in one word. On the other hand, some of the "contributing languages" do distinguish between two or among three or four of these senses. The English 'before' can, with a little good will, be used in all senses except the spatial adverbial. In Italian, though a rigorous division is maintained among 'davanti a' (sp. prep.), 'prima di' (temp. prep.), 'avanti' (sp. adv.), and 'prima' (temp. adv.). In the englishing or italianating of Interlingua, then, the clues for the correct translation of 'ante' must be gleaned from the syntactical structure of the sentence and from the semantical context of the discussion. The former sort of clue should tell whether 'ante' is an adverb or a preposition; the latter sort should tell whether it is used spatially or temporally. This kind of analysis could be avoided altogether, for 'ante' anyway, if Interlingua itself used four different words instead of the single word 'ante.' The Italian words might profitably be taken over here by Interlingua, with the insertion of a hyphen in 'prima di' so that it becomes 'prima-di' (or 'prima-de'), and with the elimination of the unattached 'a' of 'davanti a.' Just as it simplifies the interpretation of idioms and reflexive verbs to hyphenate or otherwise to agglutinate them, there is no logical reason why an adverb or a preposition should consist of several disconnected words. English, incidentally, is not entirely free of such illogicalities. We say 'near the barn,' but 'far from the barn;' 'behind the table,' but 'in front of the table.' In treading among the Interlingua particle system in search of ways to improve the language's rote translatability, we must of course awaken no more sleeping dogs than necessary. To some extent, the asseveration that Interlingua can serve as an intermediate language conflicts with the more frequent claim that it is an easily read and easily learned auxiliary tongue. If we attempt to make it more logical, we may in so doing render it less readily comprehensible. (A good example of this is the artificial language "Loglan" of James Cooke Brown, as described in his article, "Loglan," *Scientific American*, June, 1960). The modifications of Interlingua that we suggest are not *in toto* so far-reaching that they should make it harder to read or to learn. It may be more of a bore to

learn four words than one, as in the case of 'ante,' but the precise indication of idioms and reflexive verbs should if anything make the language easier to read. Generally speaking, any modification that improves its rote translatability should also improve its legibility, for the reason that we ordinarily read a foreign language not perfectly familiar to us in a word-by-word fashion anyway. Only when we get bogged down in our word-by-word scanning do we contemplate the possible presence of idioms, reflexive verbs, multiple meanings, and what not.

In revising the Interlingual particle system we should be guided by the general principle that two or more "important" (a hard word to define in this context) senses should not be confounded in the same word. Pragmatically, a distinction may be considered "important" if it is drawn in one or more of the "contributing languages" into which we would like to translate. Some of the "important" distinctions, then, will be spatial v. temporal, adverbial v. prepositional, adverbial v. adjectival, and other distinctions between parts of speech. (If we were devising a more rigorously logical artificial language, we might decide that some of these distinctions were unnecessary.) Others will be distinctions among various spatial relations, e.g. above v. below, and among various temporal relations, e.g. before v. after. It will not be necessary without to distinguish two meanings of 'or,' the inclusive and the exclusive, corresponding to the Latin 'vel' and 'aut,' since of the contributing languages only Latin insists on this, and few if any people are interested in the mechanical latinisation of Interlingua.

With the foregoing remarks in mind, we may next consider some of the more confounding Interlingual particles, and perhaps revise or restrict their meaning to some extent.

The primary meaning of the preposition 'de' is 'of,' in the sense of 'belonging to' or 'pertaining to.' Hence, we may restrict 'de' to this one sense, and use other words for the other senses, as follows:

belonging (or pertaining) to = de
 by means of = per-medio-de
 from = ab
 made of = fato-de
 since (temp. prep.) = desde
 with = con

The prime meaning of 'super' is the spatial preposition 'over.' Thus, we have:

about (i.e. anent) = re
 above (sp. adv.) = in-alto
 concerning = re
 on (sp. prep.) = sur
 on top (of) = sur
 over (sp. prep.) = super
 upon (sp. prep.) = sur

The word 'que' occurs in at least two idioms, 'a-menos-que' ('unless') and 'lo-que' ('that which'). These will

cause no trouble so long as they are hyphenated or compounded. Outside of these contexts its primary sense is the relative pronoun and conjunction 'that.' Thus, we have:

than (comp) = che
 that (rel. pron., conj.) = que
 that which = lo-que
 what (interr. pron.) = qual
 what? = come?
 which (interr. pron.) = qual
 who (rel. pron.) = qui
 who (interr. pron.) = chi
 who? = chi?
 whom = chi

We may analyse 'per' as follows:

by (for passive constructions) = per
 by means of = per-medio-de
 during = durante
 for = pro
 through (sp. prep.) = a-transverso-de*
 through (sp. adv.) = a-transverso
 throughout (temp. prep.) = durante

Compounds of 'per,' 'pro,' and 'que' include 'perque' and 'proque.' To avoid ambiguity, we suggest using 'perque' in the sense of 'because' and 'proque' in the sense of 'why?'

We may analyse 'si' as follows:

if = si
 so (adv.) = sic
 so (comp.) = cosi
 yes = oui

For 'como,' we have:

as = como
 how? = come?
 what? = come?

For 'isto:'

this (pron.) = isto
 this (dem. adj.) = iste
 these (pron.) = istos
 these (dem. adj.) = istes

For 'omne:'

all (adj.) = omne
 all (pron.) = totes
 all the world = toto-le-mundo
 each = ogni
 everyone = totes, tutti
 everything = toto, tutto

* There is no exact interlinguicism for 'through' in the context of such phrasal verbs as 'to see it through' and 'to muddle through.' These and similar phenomena are essentially local from the point of view of "standard average European", they do not belong to the "intersection" of the important western European languages, and their meaning is only very roughly approximated in Interlingua.

To make any changes in Interlingua other than of the foregoing sort would probably be to pass the point of diminishing returns. For an infinitive like 'finir' in our earlier example, which could theoretically be translated into English either as an infinitive or as a substantive, it should not be necessary to add a separate gerundial form to Interlingua. We may reasonably suppose that a recognition routine could be devised for Interlingua that could tell when 'finir' is used verbally and when it is used substantively. In our example, the fact that 'finir' is immediately preceded by the definite article 'le' is sufficient indication that it is used as a noun. It would moreover be a shame to damage the verbal simplicity of Interlingua by bringing conjugations back in, and mechanical translation out of Interlingua does not require this. In our example, person and number for all verbs are sufficiently indicated by their directly preceding nouns or pronouns; 'grande potentias,' 'illes,' and 'paises plus parve' all require a third-person-plural form. Finally, we shall propose no changes in the word order of Interlingua, nor any routine that automatically rearranges the words into a more English pattern. English and Interlingua word-orders are sufficiently alike so that their differences alone should not interfere with the easy editability of a rote translation, and it would moreover be difficult to devise a rule, for example, that would be entirely correct for the order of nouns and adjectives. The normal Interlingua adjectival position is after the noun, but there are plenty of exceptions, and the usual English scheme of adjective followed by noun is likewise exceptional. We shall be satisfied if we can produce a readily redactable translation of an Interlingua text, and we suggest that this is possible, assuming that some changes of the above sort are made in Interlingua. Let us examine this proposition in terms of our earlier example. According to our suggestions, it will have to be rewritten as follows:

A-menos-que le grande potentias vole-dicer lo-que illes dice re le finir del provas nuclear, multe paises plus parve s'espaventara.

If we assume the existence of a routine sagacious enough to recognise that all the verbs are third-person-plural, that 'illes' is 'they' rather than 'them;' that 'finir' is substantive, and that 'paises' requires 'many' rather than 'much,' a rote translation of the passage yields:

UNLESS + THE + GREAT + POWERS + MEAN + WHAT + THEY + SAY/TELL + ABOUT + THE + ENDING/FINISHING + OF + THE + PROOFS/TESTS/TRIALS + NUCLEAR + , + MANY + COUNTRIES/LANDS + MORE/PLUS + LITTLE/SMALL + WILL + BE + FRIGHTENED.

The only multiple choices that remain are those for 'dice,' 'finir,' 'provas,' 'paises,' 'plus,' and 'parve.' In each case here it is a matter of choosing between or among words that are more or less synonymous, and it is probably not wise to try to eliminate these choices. To list just one choice in each case would be arbitrary,

and to decide between or among them mechanically would require an extremely sophisticated routine. If all the editor has to do, is to make choices of this sort and to make some minor changes in word-order, we may safely say that the translation is "easily editable." We may next assay the translation of two Interlingua sentences taken from actual texts, for each giving (1) the original Interlingua passage, (2) the revised Interlingua passage, (3) the rote translation of (2), and (4) a correct idiomatic English translation.

1. De un latere esseva le latinistas traditional qui se monstrava preoccupate del problema de revitalisar le studios classic . . . (*Novas de Interlingua*, Vol. 3, No. 1, Jan-Feb., 1958, pp. 1-2).
2. De-un-latere esseva le latinistas traditional qui se monstrava preoccupate per le problema de revitalisar le studios-classic . . .
3. On one side were the latinists traditional who showed themselves preoccupied by the problem of revitalising the classical studies . . .
4. On one side there were the traditional latinists who were preoccupied with the problem of revitalising classical studies . . .

In this example, the hyphenating of the idiomatic and reflexive constructions 'de-un-latere,' 'se-monstrava,' and 'studios-classic' substantially improves their rote translatability. The transition from (2) to (3) presupposes moreover a routine that can recognize the plural intention of 'esseva' and 'se-monstrava' (the sole clue for which is the plural ending of 'latinistas'), that can recognize the nominative intention of 'qui,' and that can recognize the gerundial intention of 'revitalisar.' In rewriting the original passage (1) it was also necessary to replace 'del' with 'per le,' so that the meaning 'by the' would unambiguously come forth (some editors would no doubt prefer to change 'by' to 'with' in the final redaction, as we have done). Our second example is:

1. De tempore a tempore, e a intervallos progressive-mente decrescente, nos ha trovate nos embarassate per le requesta de recommendar un bon summario historic e actual del problema del communication translingual e de su possibile (o imaginabile) solutiones (*Novas de Interlingua*, Vol. 3, No. 3, May-June, 1958, p. 1).
2. De-tempore-a-tempore, e a intervallos progressive-mente decrescente, nos ha trovate-nos embarassate per le requesta de recommendar un bon summario historic e contemporanee del problema del communication translingual e de su possibile (o imaginabile) solutiones.
3. From time to time, and at intervals progressively decreasing, we have been embarrassed by the request of to recommend a good summary historical and contemporary of the problem of the communication translingual and of her/his/its possible (or imaginable) solutions.

4. From time to time, and at progressively decreasing intervals, we have been embarrassed by the request to recommend a good historical and contemporary summary of the translanguing communication problem and of its possible (or imaginable) solutions.

In going from (1) to (2) we treat 'de-tempore-a-tempore' and 'trovate-nos' as idioms. A routine that can recognize the nominative intention of 'nos' is presupposed. The adjective 'actual' has too many different English meanings, and is replaced by the more precise 'contemporary' (or 'contemporanee'). The only multiple choice word that remains is 'su,' and we'll not assume a routine sapientipotent enough to choose among 'her,' 'his,' and 'its' in all contexts.

The final question we shall raise is, just how important is it to translate from Interlingua into English or other natural languages? At present most of the journals that use Interlingua are written primarily in English, and use Interlingua only for summaries. There are only two journals, *Spectroscopia Molecular* and

Novas de Interlingua, written exclusive in Interlingua, and there are several non-English medical journals that use Interlingua for summaries. These latter include *Giornale Italiano di Chemioterapia*, *Haematologica Polonica*, *Revista Cubana de Cardiologia*, and *Archivos Peruanos de Patologia y Clinica*. If the number of non-English journals using Interlingua were to increase severalfold, and if Interlingua were to prove not readily legible by monolingual English speakers (there is some evidence that this is the case), then there would be some advantage in translating it efficiently and perhaps mechanically into English. More useful of course, would be a program that translated mechanically from English into Interlingua, or even that produced Interlingual summaries of English articles. But it is unfortunately not much simpler in principle to translate mechanically from English into Interlingua than into French or Italian, since the primary problem in each case is the unsolved one of automatically recognizing the syntactic and semantic structure of the English sentence.

Received April 1, 1961