Model English for Mechanical Translation

an example of a national language regularized for

electronic translators

(This paper be entirely writed in Model English)

For the Conference on Mechanical Translation

at the Massachusetts Institute of Technology, Cambridge, June 18-

20, 1952

by

Stuart C. Dodd

Washington Public Opinion Laboratory

Contents

Model English

an example of a national language regularized for machine translation

I.   <u>Rules for</u> <u>model languages</u>
      This paper will describe some of the rules for a model language for purposes
of machine translation.[1] English will be the example in applying the rules. The
purpose of this set of rules be to enable an electronic machine to swiftly
translate <u>from</u> model English as a "foreign" language or, more importantly, <u>into</u>
model English as a "target" language.
      A "model" language will mean here any language in which the rules for
regularity and familiarity have been jointly maximized. This two classes of rules
are here operationally defined by the ten ideal rules below. This set of rules
tentatively specify our theory of model languages:
      <u>Familiarity</u> rules for a model language:
      1. It should include the <u>most users</u>, i.e., it should maximize the
          number of persons who can use it.
      2. It should include the <u>most used words</u>, i.e., the words, etc., of
          most frequency in the standard form of the language.
      3. It should include the <u>most useful words</u>, - it should include most
          of the meanings (referents) of the standard language.
      <u>Regularity</u>   rules for changing a folk language:
      4. Every word should have a fixed position (i.e., one sequence in a
          sentence).
      5. Every word should have just one <u>meaning</u> (as an ideal seldom
          realized).
      6. Every word should have just one grammatical <u>form</u> (i.e.,
          uninflected particles).   .   .
      7. Every word should have just one <u>pronunciation</u> (the dominant one).
      8. Every word should have just one <u>spelling</u> (preferably phonetic).
      9. Every letter should have just one <u>shape</u> (such as the lower case
          printed shape).
      10. Every letter should have just one <u>sound</u> (such as in the
          international phonetic alphabet).
This ten rules roughly spell out the semantic ideal of "one-word-one
meaning" or maximum one-to-one correspondence of a symbol and its
referent. Only rules No. 4-*6* calling for just one word order, meaning, and
form will be necessary for mechanical translation, and so will be chiefly
discussed here. The three familiarity rules and the four remaining
regularity rules will be discussed elsewhere.[2]
      Each rule will be further operationally defined by a percent of
fulfillment or statistical index which can measure the degree to which
that rule or ideal did be attained to date in a specific language. Thus
model English can become 100 percent uninflected, fulfilling Rule No. 6
perfectly, by making every word an uninflected particle as below
described. Similarly, the other nine rules be maximally attainable in
model English.
      The two kinds of rules will be kept in balance. A model language
will seek the most familiarity that be possible for combining with the most
regularity. The standard forms of any national languages, such as current
English, represent the extreme of great familiarity and great irregularity.
A regular artificial language such as Esperanto (or still more our Tilp[3],
or most of all logic and mathematics itself) represent the opposite extreme
of unfamiliarity and of regularity. At the irregularity extreme a language

---

[1] For further description of model English for general international
auxiliary language use, see:
   Dodd, S. C., "Model English," Gen.Sem.Bull., No.8 and 9, Fall and
Winter; 1951. Systematic Social Science, University Bookstore, Seattle,
1947.
[2] For these indices see:
   Dodd, S.C., "On Measuring Languages." Jour.Amer.Stat.Assn., Vol44,
   Mar.1949.
[3] See Dodd, S.C., "Tilp - a ten letter alphabet of meanings," Gen.Sem.
   Bull., No. 6 and 7, Spring and summer, 1951.

may have as many rules as it have variant usages; at the regularity extreme
it might have only one rule; therefore, the number of rules needed will be
one measure of the position of a language on the continuum between the
poles of familiarity and regularity. The ten rules and they subrules below
more exactly described try to specify the balance in model English between
the familiar and the regular more exactly.

     The two classes of rules should be equally maximized both in
theory and in practice. The below rules briefly state the theory; writing
texts in model languages puts the theory into practice. To do this, we
write the text *of* this paper entirely in model English to show how the
theory of semantic models will work in practice. To further empirically
test this theory, a former paper describing model English did be writed
entirely in model English. It be being translated into German, French,
Spanish, Italian, Russian, Hebrew, Chinese, and Japanese (but not into the
model forms of these languages since model Chinese, etc., have not yet been
developed).

     We use the "regular" model English, using rules 1, 4, 5, 6, and
9 here. (This also be called "Ruly English" for "rule-obeying English," in
contrast to standard unruly English.) We sha11 not here follow rules 2 and
3 which limit the vocabulary to the few most used and most useful words,
since machine translators can handle millions of words. We shall not here
follow rules 7, 8 and 10 which call for no capital letters, phonetic
spelling and phonetic letters, since machine translators can readily handle
the present standard spelling and the English alphabet.

     Since the omitted rules decrease familiarity and since their
increase of regularity be not needed by the machines, they be not followed
here. Here we describe only "simple English" - one of the three subclasses
of model English - as follows:

          Simple (model) English - with grammar regularized (Rules 4 and 6
          obeyed);
          More Simple (model) English also have a limited vocabulary
          (Rules 1, 2, 3, 5, and 7 also obeyed)
          Most Simple (model) English also have phonetic spelling (Rules
          8, 9, and 10 also obeyed).

     A warning to the reader may be helpful here before describing the
three rules for a single word order, meaning, and form. The rules here may
need much refining by linguistic experts. The author be an amateur in this
field. He did develop model English as a by-product of dimensional
analysis in sociology, which did lead to studying semantics. Dimensional
analysis, when applied to language, did result in inventing "Tilp"[3] and the
more dilute but immediately more useful model English. But the detailed
below rules may have aspects which seem naive to a linguist. To get their
help in improving it, for possible use in mechanical translation, be a
major aim in presenting it here.

II.   Rule No. 4 – one word order

     The overall semantic rule of maximizing the one- to-one
correspondence of symbol and referent mean that the position or order of
the words in a sentence should tell whether each word be the subject or
verb or object or modifier, etc. To fix the meaning of the word order, we
first, as usual, list the variant word orders used in English and choose
the most used order so as to maximize familiarity. Then we standardize
this word order as the only one permitted in order to maximize regularity.
These twin steps seem to result in the four following rules for word order
in model English:

          4a. Put the subject first, then the verb, then the direct object,
               then the indirect object.
          4b. Put adjectives, adverbs and auxiliary verbs before whatever
               they qualify, e.g. "We will always speak this way."
          4c. Put all othe qualifiers, such as the suffix particles (es,
               er, ed, ing), phrases and clauses, after whatever they
               qualify, e.g., "The speaker who do otherwise be breaking the
               rules of model English."

4d. Put the subject after the first verb auxiliary in a question, e.g., "Did he say this?", "Do you understand?", "Will she do it?"

Amplifying these rules will be needed to cover the field more completely.

Note that a single or standardized word order in sentences will make English a more flexible language, in that a word may be used as different parts of speech - as noun, verb, adjective, etc. Thus nouns or pronouns or verbs become adjectives if put before what they qualify, e.g., "door man", "we group" (for "our group"), "fly boys." Words may be clearly used as verbs if they connect subject and object and have the auxiliaries of a verb, e.g., "He will chair the meeting.", "We will tea at four today.", "Gossip may black a reputation."

As usual we can get more regularity by forcing everything under a few rules but with great loss of familiarity. Thus rules 4b and 4c might be combined into the one rule, "Modifiers follow what they modify." But using this rule will give unfamiliar sequences like "particles, adjectives, and verbs in English put will be in order unfamiliar a very."

To measure the degree of fulfillment of this rule in any text containing N words, count the percent of words that conform to the rules of word order. By this index[4] model English can score 100 percent in following the word order rules without exceptions. This index complete the identifying operational definition of a single-word-order.

III.   Rule No. 5. - one meaning for each word

The semantic rule, "one symbol, one meaning" be an ideal which can be approached but seldom can be perfectly and permanently achieved. For words may shift their meaning and may develop multiple meanings. This may require qualifying adjectives to distinguish between meanings, as in the statement "When you say 'How funny,' do you mean 'funny ha-ha' or 'funny queer'?" Thus the new meaning of "tank," whenever the context did not make the new military meaning clear, may require a phrase like "army tank" to distinguish it from the familiar "water tank."

Of course, systematic ambiguity, multiple meanings, and highly inclusive concepts have they uses in language and literature. But for purposes of exact communication, a single well-defined meaning for any word or phrase be the ideal. To carry out this ideal, was such aids as the Basic English Dictionary, which select the chief meaning of each of 40,000 words (and define they in terms based on the 850 basic words).

For machine translation purposes, words can be coded with multiple meanings of course. But the machine must have some instruction, from a subcode or from context, as to which alternative meaning to act on. The supplemental codings in effect carry out the rule of "one symbol, one referent."

To measure the degree of fulfillment of this rule in any text will require counting the words and the meanings, to compute a "univocality index." The numerator of this index be the number of words in the text at issue times 100. The denominator be the sum of the words when each be multiplied or weighted by the "rank" of it as to multiple meanings. This rank be 1 for the first meaning of any word; the rank be 2 for each recurrence of that word with a second meaning; the rank or weighting coefficient be 3 for each occurrence of a third meaning; etc. This univocality index will be 100 percent if every word have just one meaning. The univocality will sink below 100 percent in proportion as multiple meanings recur.

This index will constitute the identifying type of operational definition of univocality for it will identify exactly how much univocality characterize a particular text. The other type of operational definition - the generative definition - be the use of the above dictionary - since this

---

[4] For these indices see:
   Dodd, S.C., "On Measuring Languages." Jour.Amer.Stat.Assn., Vol.44, Mar.1949.

can help the operator to produce a univocal text.

IV.   <u>Rule No. 6 - one form for each word - no inflections</u>.
        The semantic rule of seeking the most full one-to-one
correspondence of symbol and referent will call for the forms of grammar
to be completely analyzed into invariant particles, each with constant
meaning. This will mean that words will be not inflected by affixes or
other change of the word form. All grammar will be showed by uninflected
particles or syntax words in combination with substantive words. This
particle for any item of grammar can be selected from the great variety of
usages in standard English, so that no unfamiliar neologism will be needed.
How this will work will be spelled out below for inflections of case,
number, person, degree, tense, voice, mood, participles, nouns, and verbs.
        A. <u>Case</u>. The general rule, used throughout model English, that
words become regular, not by inventing new forms but by merely choosing the
rule-obeying forms out of the wealth of different uses in current English,
may be showed in the possessive case. We express possession in three ways:
by " 's," by "of," or by the adjective, as in "the adjective's use," "the
use of adjectives," "the adjective use." In order to replace inflections
with uninflected connecting words and name-words, we may use only the last
two ways.
        Since the direct object always follows the verb (by rule 4a), it
need no change in ending. This, in English, mean replacing the pronoun
"us" by "we"; "him" by "he"; "them" by "they"; "whom" by "who"; e.g., "I
like he." The indirect object would have the connecting word "to" in front
of it, e.g., "I will give it to she."
        B. <u>Number</u>. The plural ending might be made a separate word – the
plural particle – "es" meaning "more than one of..." in Model English.
Then as a yielding to familiar habits, joining it to a noun, as in current
speech and writing, might be permitted. Then the only unfamiliar use would
be making the irregular plurals regular, such as "child es" or "childs" for
"children", "woman es" or "womanes" for "women," "thises" for "these,"
"thats" for "those," etc.
        C. <u>Persons</u>. To get rid of inflections of person from the
 English verbs we need only to drop the unnecessary *"s"* in the third person
 of the present tense and extend one form to all persons in certain
 irregular verbs, such as "have" for "has" and "be" for "am," "are," and
 "is," e.g., "he go," "he see," "he have," "he be," etc.
        D. <u>Degree</u>. The unchanging words "more" and "most" can easily be
used for the degrees of adjectives and adverbs.
        E. <u>Tense</u>. Take the present tense as the verb and let it not be
inflected. The perfect tenses can be restated by such words as "already,"
"before that," etc., e.g., "He will come before then." for "He will have
come." The present tense would not be changed even for number or person,
and the *f*uture tense would take "will," while the past tense would put
"did" before the verb. This use of "did" for the past tense will extend
the current use for the sentences which contain "not" or a question or a
strong statement, to all sentences. Thus we will say "I did go" for "I
went" in a declarative statement much as we now say "I did not go"*,* in a
negative statement of "Did you go?" in an interrogative statement and
"But, I did go" in an emphatic statement. This regular makes every past
tense. Thus "taught" will become "did teach," "was" and "were" will
become "did be", "had" will become "did have," etc.
        F. <u>The passive</u>. The verb "be" with the passive participle "-ed"
can express the passive voice much as at present, e.g., "I be loved.", "He
will be loved.", "She did be loved."
        G. <u>Mood</u>, <u>etc</u>. The present "helping verbs" "do, say, might, let,
can, would, should, ought, must" with the uninflected main verb, can
express all moods and other shades of meaning of a verb.
        H. <u>Participles</u>. The two very useful participle endings, "ing"
and "ed," can be separate words, meaning, when standing alone, "the
acting" of the verb and the object "acted upon." Again, in order to keep
as much of the familiar as possible, this new words might be permitted to
be joined to the verb so that one can write and speak they as at present.

"Ed," the passive participle, will get rid of all that be not regular in verbs though introducing some forms which be not familiar to most people, however natural to children, such as "speaked" for "spoken," "sayed" or "sayd," for "said," "rund" for "ran," etc.

        I.. <u>Nouns from verbs, etc</u>. The partciples be verbs changed to adjectives and, like other adjectives, may be changed to nouns by putting simply "the" before them, e.g. "the loving," "the loved," "the good," "the true." This make most of the noun endings such as "tion," "ment," "ence," "ness" not needed. The verb roots of such words (adjective roots in the case of "-ness") should be in the model English list and then adjectives and nouns can be maked easy and regular by the separable words "ing," "ed," "er," and "the," e.g., "the paying" for "the payment," "the acting" for "the action," "the existing" for "the existence," "the bad" for "badness," etc.

        A fourth joinable or separable word be "er," meaning "the acter." It will change a verb into a noun of agent, e.g., "doer" or "do er," "writer," "seer," "eater."

        J. <u>Verbs from nouns or adjectives.</u> A noun or an adjective may be used as a verb if the meaning be clear, because of it position between subject and object or because of it being preceded by an auxiliary verb or being followed by a verb particle, "ing," "ed," "en," e.g., "Caves house the refugees.", "I sweeten my coffee.", "They will shrapnel that area.", etc.

        Rules such as these can make English as free of inflections as Chinese or as the most model artificial language. The four basic particles, or separate but joinable words, "er," "ing," "ed," and "es" when joined to verbs (or to nouns in the case of "es") will combine being separable words with the familiar current use. Such affixes as "un" for "not" as in "uninflected," "en" (to make, as in "sweeten"), "ess" (for the feminine as in "lion-ess"), may be called particles or separate words (which may be joined in use for familiarity). Thus the sixth rule, that every word shall have one and only one form, can be 100 percent satisfied in model English.

V. <u>Summary</u>

        For mechanical translation, a model language may be useful at the pre-editing or post-editing stages or both. Model languages here mean any national languages which have maximum regularity while keeping maximum familiarity. More exactly, a model language be operationally defined by the ten rules and subrules stated here and illustrated in the case of English. All the rules here spell out the general semantic ideal of just one meaning for any symbol. Of the ten rules, the three most needed by the machines be that words should have just one order, one meaning, and one form in grammar. Model English can achieve 100 percent regularity by standardizing: a) the chief word order; and b) the chief meaning of each word; and c) all inflections of grammar by using particles (which be invariant syntax words like "es," "er," "ed," and "ing," for plural, the acter, the acted on, and the acting, respectively.)

        Model English can also keep a high degree of familiarity by choosing the <u>rule-obeying</u> forms from the many variant usages in standard English. In fact, Model English might be more exactly called "Ruly English." We did write this paper about model English entirely in model English to show how much familiarity can be combined with complete regularity in a model language.

plate 1   **model english**   **U:52-96**

an international auxiliary language which be perfectly regular yet

familiar to several hundred million people

**let "model english" be defined by the ten rules and percents stated here
in model english**

10 rules	percents	application to english	percent
as ideals for any language for world use	for comparing languages (from a sample million sentences)	towards an ideal international auxiliary language the choice of rule-obeying parts out of the unruly usages in current english can yield a regularized or model english. it can match the most ideal artificial language in having rules with no exceptions, this model english will be large-ly familiar to a sixth of the world now and will be easy for the rest to learn.	possible in model english:

3 rules for most familiarity

1. most maximal users	% of world knowing each word, averaged	adopt the most international words from english as the most widespread, influential, and fast-growing of world languages.	
2. most maximal used	% of first 1000 words in sample above	grade words into lists by they frequency of use. say what level be used, adopt the lists of faucet, palmer, thorndike and west. call most used 1000 words level a. call most used 2000 words level b. etc. •	
3. *most* 100% useful	% of basic english words	adopt in level a the 850 word list of basic english as the most useful set of less than 1000 words	

7 rules for most regularity

4. sentences 100% should have one order	% of words obeying rules	a. put subject, verb, object, indirect object in this standard order. b. put adjectives and adverbs and helping verbs before whatever they qualify eg. "we will always speak this way." c. put other qualifiers such as phrases and clauses after whatever they qualify. d. put subjects after the first verb particle in questions, eg. "do we not?" "will he go?" etc.	
5. words 100% should have just one meaning	% of words with chief meaning	extend the "general basic english dictionary" which tell the chief sense of each of 20,000 words in terms of 850 words. restate idioms till each word be translatable. ' eg. "to put up with" = "to bear", "get under weigh"= "start moving".	
6. one form 100% (i.e. no inflections)	% of words not inflected	use particles instead of inflections, as follows: a case: "of", not "s" for possession; eg "of men", for "men's" direct objects follow verbs (rule "of i, he", etc. for 4c); "to" before indirect objects. "my, his" eg i give he to she" b. number: "es" "more than one of" = eg. "ones", "word es" (or a joinable particle for all plurals. "words"), "mens", "thises", "whats"	

10 rules	percents	application to english	percent
one form (i. e. no in- flections	% of words not inflected	c. gender: none, the "ess" means "female"	eg "lion ess" or "lioness"
		d person: no verb inflections for person "be" for "am", "are", "is".	eg "see" for "sees", etc.
		e adverbs: the joinable particle "ly" (=like) may be omitted	eg "write sensible" or "write sensibly"
		f degree: "more" and "most" for "_er" "_est"	eg "more good" for "better" etc.
		g tense: "did", "will" for past, future. -	eg "did be" for "was, were" "did go" for "went" etc.
		perfect tense use "before then" passive	eg "he did go before then" for "he had gone".
		h, voice: "be" and the passive participle	eg "this be writed by rule 6h"
		i. mood: "to" (infinitive), etc.'do' (strength), "may" (possibility), "let" (permission), "con" (ability), "should" (weak duty), "ought" (strong duty), "must" (necessity),	eg "if i do try to speak model english to he i con do it, altho i may fail at first. ' let i try, i ought succeed".
		j participles: the joinable active participle "ing"= "acting", the joinable passive particle "ed" = "acted on"	eg "all verbs be maked regular with unchanging roots by rules 6 g - j"
		k nouns: "the" for "ness" "ity"; "the" and "ing" for "tion", "ment", "ence".	eg "the good, the pure" for "goodness, purity", etc. "relating" for "re-lotion" "paying" for "payment", etc.
		l "er" = a joinable particle of "agent"	eg "payer, or pay er, re later" .
7. one pronun- ciation	% of words with chief pronun- ciation	adopt most heard pronunciation as official.	100%
8. one spelling	% of words spelled as pronounced	expect from phonetic spelling: at first, unfamiliarity and op-position from present readers; later, more people able to read and with less schooling .	100%
9. letters should have one shape	% of letters with chief shape	use the small printed latin letters for handwriting and for all capitals, eg. "model english"	100%
10. one sound	% of letters with chief sound	pronounce long (:) and short u as in "truthful" " " " " a " " "yardman" " " " " i " " "shielding" " " " " ɘ " " "earnest" " " " " o " " "sort of" " dipthongs as in so;und, o:l, ve:il. omit the 5 long vowel marks (:) when learned.	100%
ideal = 100%	weighted over-age %" = "per-cent of per-fection"	write words as they sound as below: yo:z the ten persents to: komper langwijez	maximal
		a) %#, the mo:u:st important ov o:l, wil sho:u artfishol langwijez az li:st familyar and inglish az mo:u:st familyor:	
		b) other persents will sho:u: artifishol langwijez as mo:u:st regyular and inglish az li:st regyular.	
		c) o:l persents will sho:u: that model inglish kan bi: the mo:u:st perfekt ov a:l langwij.	