

ON PROBLEMS OF ADDRESS IN AN AUTOMATIC DICTIONARY OF FRENCH

by

P. MEILE

(Ecole des Langues Orientales, Paris)

1.1 In most printed dictionaries, the address of each article, that is of each set of information pertaining to that particular entry, is simply the word itself. It has to be so in a book for common use: for the general reader's sake, the word must be entered in its complete form.

2.1 In the case of long words, part only of the letters contained in the word would be enough to provide an adequate address, that is to achieve an alphabetical classification. As a matter of fact, the last letters of a long word (say a word of more than ten letters) do not play any part whatsoever as classifiers. The first four or five letters are very often sufficient; subsequent letters provide an over-definition which, from the point of view of address only, remains useless.

1.3 A mechanical address is not subject to the same conditions as an address in a printed dictionary. Letters which are meant for a human reader only and are not indispensable for purposes of classification, become completely superfluous. Pronunciation, visual "recognition", aesthetics, etymology, etc. are no longer relevant, and only classification matters.

1.4 For classification purposes, combinations of letters are available, just as in a printed dictionary. But, in a machine, we have at our disposal further information which is not made use of in a printed dictionary, and which, so far, seems to have been neglected, generally speaking, i.e. the number of letters of the word. As a matter of fact, "wordlength", expressed with accuracy as the number of the letters that make up a word, is an important feature of the word. And it so happens that counting, a task so tedious to Man, is just the thing machines can do.

1.5 Classifications can also be based on other data, such as categorization of the word, its construction, etc. To the extent that such means of classification are external ones, they must be used with moderation; but to the extent that final letters of a word are grammatical marks, from which

a categorization is apt to be inferred, such final letters are characteristics properly belonging to the word itself, and we are entitled to use them with a double purpose: (1) address, (2) analysis of the function of the word. Characteristic letters are to be looked for in the end of the word as much as in the beginning. This, of course, applies to the French language.

1.6 Under such circumstances, it seems that an original conception of a mechanical address would consist in describing a word by its first and last letters, together with the indication of the total number of letters in the word. The present paper attempts, by a trial and error application of such methods, to reach a general formulation expressing the most economical address within a given glossary. But such a formulation is a target for a remote future, and, in the following pages, we shall content ourselves with a few examples.

2.1 Let us suppose, to begin with, a classification by four initial letters. In a French dictionary, the four letters m é n a begin the following words:

ménade	ménagère
ménage	ménagerie
ménagement	ménageur
1. ménager; 2. ménager	

There is no other word, provided, of course that due account is being taken of accented e, written as é.

Enumeration of letters gives the following classes:

6 letters: ménade, ménage
7 letters: ménager (1 and 2)
8 letters: ménagère, ménageur
9 letters: ménagerie
10 letters: ménagement.

Ambiguities on six and eight letters consist in pairs of words differing widely one from another in frequency: ménade (6) and ménageur (8) are words of a particular rarity and can safely be omitted in a current glossary. Ambiguities are thus disposed of by omission of one of the constituents.

In this state of things, if the tetragram ména is followed by the figure of the letter count, we have:

ména 6 = ménage
ména 7 = ménager (1 and 2)
ména 8 = ménagère
ména 9 = ménagerie (menageuse is discarded)
ména 10 = ménagement

2.2 It will be noticed that an ambiguity remains about ménager, which can be both a verb and an adjective. This is no matter of surprise or disappointment: such a polysemy does occur in the word itself, and nothing has been changed while we were operating on the address. The ambiguity cannot be solved unless grammatical characteristics are added:

ménager transitive verb
ménager adjective

We may write as well:

ména 7 v
ména 7 a

Practically speaking, ménager as a verb is often preceded by à: à ménager. As an adjective, it is never preceded by à, nor is it preceded by an article: in case of the definite article being used, ménager is a verb.

We might go further into this analysis: but it has already suggested that external criteria might be added to an abridged address just in the same way as to the word in its full length.

2.3 In the instance given above we have availed ourselves of initial letters only, and we have shown that, when combined with a letter-count, they provide an adequate classification. But final letters contain useful information concerning the function of the word. In the present case, there is obviously advantage to be drawn from the final -ment, -erie, -gère, in the way of grammatical analysis.

We are thus led to bring in the four finals, besides the four (or

three) initials. That is to say that a seven letter word will not be abridged. Abbreviation is contemplated for words of eight letters onwards. From ten letters onwards, abbreviating means a noticeable saving.

2.4 We should not lose sight of the fact, that, where a verb is concerned we shall have to trace not only the infinitive form as it stands in the glossary, but also conjugated forms. And for such forms, four finals seem indispensable. In some cases they are utterly inadequate (for instance, in an ending such as -raient, conditional, 3rd person plural).

2.5 But a further saving is probably to be effected through symbolising the finals. For, in French, word ends are very remarkable for the restricted use they make of the alphabet. Possible word ends are limited in number, and are thus liable to be represented through a much simpler system than four letters picked out of an alphabet of thirty-nine.

3.1 In the French dictionary, with the exception of proper nouns; some letters never occur in word-ends; a good many are very rare at that place. Out of thirty-nine signs which are usually needed to write French, about a dozen only are frequent in word-ends.

The following letters are either absent or very rare in word-ends:

à â b c ç è ë ê g h i î j k m o ô p q ù
û û v w y

amounting to twenty-five signs.

It is noteworthy that those letters have two characteristics: (1) rarity as last letter of a word; (2) absence of semiological value. Even the most frequent ones, -c and -o, each of which is final in about one hundred words, do not play any morphological role. As a counterpart, among the same letters are found finals of proper nouns: -c, -cq, -m, -o, -y. Many place-names end in -ac or in -ay.

3.2 If the twenty-five signs listed above are deducted from the total of thirty-nine, we have a remainder of fourteen signs with two characteristics: first, they are frequent as finals, second, they fulfill a semiological role.

Letter d is relatively rare as a word-end; it occurs in about two hundred words. Then come f and l, which characterise for instance adjectives.

Final i is rather frequent.

The following eleven letters are important as finals:

a e é i n r s t u x z.

-n and -x belong to nouns.

-z is almost exclusively verbal: it occurs as the final of the 2nd person plural. There are very few nouns ending in -z. For instance, if, in a given text, the 2nd person plural is not used, the frequency of z drops to an insignificant value, not only as final, but all places included.

Letters i, u and r are often finals of verbal forms: -i and -u commonly denote past particles; -r is mainly an infinitive ending.

3.3 Letters e, -é, -s, and -t are those conveying the least specific information, or, to put it differently, those which totalise the maximum number of values. With -s and -t, if we are to solve ambiguities, we have to bring in not only the penultimate, but also the antepenultimate letter: three letters are thus indispensable, and the fourth letter may have to be brought in as well. For instance, group -nt has such a high frequency that the antepenultimate has to be brought in, in order to get a workable sorting of -ant, -ent, -ont, to quote only the main classes. Moreover -ent is still too rich to avoid bringing in the fourth letter (i.e. the fourth when counting from the end): with the fourth letter it will be possible to distinguish, on one hand, the host - a very large one Indeed - of words ending in -ment, and, on the other hand, those which end in -dent, -lent, -nent, etc.

3.4 Restrictions in use, which are, as we have just seen, so remarkable on finals, are found also on the penultimates (i.e. letters in the second place, counting backwards). The thirty-nine signs of French are far from occurring uniformly at that place.

Researches on this particular point are still in process, and will be treated in an **Appendix**. As a first estimate, however, digrams made of penultimates and finals can be evaluated at a figure below 150, under normal and current circumstances: it seems that two letter combinations in word-ends do not amount to more than 150; and a part only of this total are combinations with a strictly morphological value.

Such a limitation is due to restrictions both in the set used for

the last and penultimate letters and in their succession. Let us mention, for instance, the paucity of words in -or or in -oc (none of them of more than seven letters), the absence of i, ô or ù as penultimates.

3.5 In **Appendix I** a study has been made of a particular case, namely of words ending in -bant. It comes out of our scrutiny that the forty-two French words ending in -bant, classified in a first instance according to letter-count (words of 6, 7 letters, and so on) can then, as a second step, be isolated through initial diagrams, or by combination of letter 1 and letter 3. Bringing in all three initials is not strictly compulsory. Provided the lexicon entered into the memory contains only usual words, the abridged address is satisfactory.

In such a case, the process means an economy for every word over eight letters, and even for eight letter words.

3.6 As a matter of fact, the economy amounts to more than that, for we have simultaneously identified a suffix with a semiological value, -ant being a present participle's ending: the grammatical function of the word is supplied by the address itself. Assuming as a hypothesis, seemingly not an unreasonable one, that we finally symbolise trigram -ant in a code which might be, let us say, of nine options, the address might be formulated as follows:

letter 1 - letter 2 (or 3) - total number of letters - 4th letter counting from the end - -ant suffix, or: 1/32, 1/36, 1/16, 1/36, 1/500.

At the beginning of the word, letters 1 and 3 give better results than 1 and 2. For figures 1/32 and 1/36, see **Appendices**. Figure 500 as the total of final trigrams is merely a hypothetical value, supposing the case of a limited dictionary.

As for the total number of letters, we have written it as 1/16, that is one figure in a key-board of sixteen, because abbreviations will work from eight letters on: thus 16 will symbolise words of 23 letters onwards: there are very few words of 23 letters or more.

The advantage thus gained is not only one of economy, it lies also in the constancy of word length, which is an appreciable feature when machines are concerned.

4.1. It is to be hoped that research be conducted, both manually and mechanically, on the lines of our sample-study on -bant. Nothing more than a working hypothesis is to be found here; we have to await the results of experiments.

4.2 The main result to be expected from experiments which might be undertaken would be to define a relation between the richness of a glossary and the quota of information required for the address therein. Language being a field with so many intricacies, such a relation will be defined exclusively on an experimental basis, even in face of studies on letter distribution in French, of the kind produced by M. Rene Moreau, which allow a certain degree of prevision.

4.3 It is obvious, from the very outset, that, in a limited glossary, for instance, in the vocabulary of fundamental French, the address can be drastically abridged. As the lexicon grows, the address will naturally develop accordingly; but sudden spurts may happen in the process of enlargement and, in the face of such events, one can imagine that a control be set upon each new entry, with a view to estimating if the cost required by a new entry is not disproportionate with the value of information such a word carries with it.

For instance, in the case of -bant ending, it looks very likely that it would pay to prevent cohobant, an extremely rare word, from entering the glossary, rather than to shift, for the sake of only one word, from initial digram to trigram.

4.4 In the foregoing remarks, reference was made only to the French language. The situation is different if various languages are considered. Let us just mention that what has been said of classification of word endings can apply even better to Russian, where classification of suffixes is an indispensable requirement in analysing the function of the word. A difficulty might arise, as it seems, in Russian, as regards prefixes, which interfere with recognition of initials. The method suggested in this paper cannot apply, as a whole, to various languages, including Russian, unless thoroughly readapted.

APPENDIX I

French words ending in -bant

The principles explained above have been verified on a sample: that of French words ending in -bant.

The ending -bant was chosen as a particular case of the ending -nt, which has a high frequency of occurrence and fairly wide uses, so that it performs an important morphological role. The ending -ant, also, is very frequent, being the ending of all present participles.

The word-ending -bant comprises an element -ant which is morphemic, and a consonant of lexical classification - b -: it derives, save one exception, from verbs whose root has a final -b.

The rhyming dictionary lists 42 words in -bant: 41 of these words are present participles, as there exist in French 41 verbs in -ber; the 42nd word in -bant is the adjective probant, which is equivalent to the present participle of the verb prouver.

These 42 words in -bant are divided up in the following manner:

3 of 6 letters	13 of 8	1 of 11
6 Of 7	12 Of 9	1 Of 13
	6 Of 10	

These words are analysed in the tables which follow. Those of 6 and 7 letters are classified by their initial letter only. In the hypothesis in which we wish to use one only of the first three letters, it is the first letter which will be the best to use, for the second and third letters give, each one individually, binary groupings; and even gerbant/herbant cannot be distinguished by the second and third letters combined, whilst they can by their initial letters. But words of six and seven letters are only included here for completeness, as only words of eight letters or more are accorded an abbreviated address.

Considering first those words of eleven and thirteen letters, which give rise to no ambiguity, they are excellent examples of the advantages of the proposed method since their address can be reduced to the combination of a suffix, with the number of letters. The initial trigram (sur- and des-, respectively) is superfluous; however, it will be recorded for reasons of uniformity.

It is the words of eight nine and ten letters which must be given most attention. These three lists show certain groupings which comprise two, three and even five words. These groups are resolved, for the most part, by the initial trigram, and even by two letters only, these being letters I and II, or I and III. The difficulties arise either from very rare words, which one can eliminate without inconvenience, or from prefixes: a word like engerbant is grouped, because of its prefix, with englobant, and recourse to the letter IV is inevitable, if one cannot have the solution of eliminating engerbant as a very rare word and little used. A word like cohobant, which requires three letters for its classification, can be eliminated, for it represents practically nothing.

On the whole, one can see that, in the example considered, and having eliminated several rare words, classification can be made in the following manner:-

- (1) suffix of four letters,
- (2) number of letters in the word,
- (3) two letters, which can be either I and II, or I and III

However, the list of words in -bant is relatively short; in many other cases, it will be necessary to take three initial classification letters; either the initial trigram or the letters I, III and IV. And the problem posed by prefixes, of two or three letters or even more, remains difficult.

On the other hand, if we consider the words in a text, with verbal or nominal endings, we will find sequences of four, and even five, six and seven final letters which carry no indication of lexical order; for instance:

commenc -eraient

The method proposed, in the form studied here, is better suited to a classification in the dictionary than in texts. It will have to be completed to be usable at all.

1. WORDS ENDING IN -BANT (TNAB) N < 8

Classification on the three initial letters

(a) Words of 6 letters

Total: 3 words.

LETTER I	LETTER II	LETTER III
C ubant	g O bant	Included in the final tetragram.
G obant	c U bant } c/t	
T ubant	t U bant }	

(b) Words of 7 letters

Total: 6 words

LETTER I	LETTER II	LETTER III
B ombant	d A ubant	bo M bant } b/t
D aubant	g E rbant } g/h	to M bant }
G erbant	h E rbant }	pr O bant
H erbant	b O mbant } b/t	ge R bant } g/h
P robant	t O mbant }	he R bant }
T ombant	p R obant	da U bant
No group	2 groups, resolved by letter I	2 groups, resolved by letter I

The word herbant is rare, and can be eliminated from the lexicon.
The word gerbant also, if need be, can be eliminated.

In this series of words of 7 letters, either of the letters II or III give the same discrimination.

2. WORDS ENDING IN - bant (TNAB) n = 8

Classification on the three initial letters

Total of words of 8 letters: 13

LETTER I	LETTER II	LETTER III
A doubant	é B arbant	fl A mbant } f/é
C ohobant } C ourbant }	a D oubant	éb A rbant }
D érobant	d E robant	im B ibant
É barbant } É herbant }	é H erbant	eh E rbant
E xhibant	t I tubant	co H obant } c/e
	f L ambant } f/p	ex H ibant }
	p L ombant }	in H ibant } e/i

(98026)

389

LETTER I	LETTER II	LETTER III
F lambant } F ourbant }	i M bibant i N hibant	pl O mbant } p/s ad O ubant }
10 I mbibant } I nhibant } P lombant T itubant	c O hobant } c O urbant } f O urbant } c/f e X hibant	de R obant ti T ubant co U rbant } c/f fo U rbant }
4 binary groups Resolved: <u>cohobant</u> by letter III, the others by letter II	1 binary group 1 ternary group The binary group is re- solved by letter I; in the ternary group, <u>cohobant</u> is awkward	3 binary group 1 ternary group All resolved by letter I

The word cohobant, which gives difficulty, is an extremely rare word, unknown to all those French people I have been able to question about it. It is a word from the alchemist's vocabulary. It can be eliminated from our lexicon without inconvenience. After this elimination, there is not a single ternary group.

Equally rare also, and also will be eliminated: adoubant and fourbant, which avoids two ambiguities, and eherbant. There thus remain, finally, nine words which are classified, except for imbibant/inhibant on the first letter alone. The ambiguity of imbibant is resolved equally well by either letter II or letter III.

3. WORDS ENDING IN - BANT (TNAB) n = 9

Classification on the three initial letters

Total of words of 9 letters = 12

LETTER I	LETTER II	LETTER III
A bsorbant	r A doubant	in C ombant
E ngerbant	a B sorbant	ra D oubant
nglobant		
nnerbant	r E gimbant	en G erbant } en G lobant } re G imbant }
njambant	tombant	
(96026)	390	

LETTER I	LETTER II	LETTER III
P rohibant	e N gerbant	en J ambant
	globant	
R adoubant	herbant	
	jambant	sy L labant
egimbant	jambant	
10 é sorbant	i N combant	pr O hibant
etombant	p R ohibant	ab S orbant } a/r
	s Y llabant	ré S orbant }
S yllabant		
2 quaternary groups.	1 binary group	1 binary group,
(1 ternary if	1 quinary group	resolved by letter I
é is distinct	1 ternary group.	

In this series, the groups are formed by prefixed words: the prefixes en- and re- play a disturbing role, making the classification less fine.

These prefixes being, here, digrams, letter II supplies nothing more in principle than letter I. The confusion could be much greater if one did not have the help given by the fact that the prefix re- also occurs in the forms ra- and re- and that the prefix en- takes sometimes the form em-.

Where there are prefixes of digrammatic form, it is necessary to refer to letter III: thus the ambiguity of regimbant/retombant is resolved by the third letter. The same with englobant/enjambant.

There remains one ambiguity: to decide between engerbant/englobant, it is necessary to refer to the letter IV. However, we can avoid this ambiguity by eliminating the word engerbant, which is very rare. After this elimination, the classification is satisfactory on three letters. One can even be satisfied with letters I and II.

Finally, we could eliminate, as well as engerbant, the following rare words:

enherbant, regimbant

These three words removed, there remain a total of nine words which, on the letter III only, give rise to but one ambiguity: absorbant/résorbant. The letter I is sufficient to clear this ambiguity.

Difficulties of Prefixes

We can see the difficulty caused, in a method of abbreviations, by prefixes which are digrams. The difficulty is still greater when it arises from trigram prefixes.

com-, des-, dis-, par-, per-, pro-, sou-, sub-. Let us also add pour- and trans-.

Thus combattre and commettre can only be distinguished by letters IV and V: words of nine letters cannot be abbreviated according to the proposed method if one restricts it to the three initial letters. The same goes for percutant/permutant, soutenant/souvenant.

In that case, one can work in two ways:

- either refer continually to the fourth letter,
- or identify the prefixes, which are few in number and isolate them from the rest of the word: but this is a complication, the advantages of which are none too clear. Whereas the identification of suffixes is a profitable operation, that of prefixes is not very instructive.

4. WORDS ENDING IN - BANT (TNAB) $n \geq 10$

Classification on the three initial letters

(a) Words of 10 letters -

Total: 6 words

LETTER I	LETTER II	LETTER III
D ébourbant } éplombant } ésherbant }	r E courbant d E bourbant } plombant } sherbant }	dé B ourbant } d/e em B ourbant }
E mbourbant P ecourbaut S uccombant	e N bourbant s U combant	su C combant } s/r re C ourbant }
1 ternary group	1 ternary group (the same one as for letter I)	2 binary groups resolved by letter I

Here, as above, the prefixes are a disturbing element: the prefix de- is the cause of a ternary group which can only be resolved by the letter III.

A very rare word, to be eliminated: debourbant.

(b) *Words of II letters*

1 word: S urplombant

(c) *Words of 13 letters*

1 word: D ésembourbant

APPENDIX II

In order to evaluate the number of words likely to be concerned in the system of abbreviation described above, it is useful to present some figures relating to the distribution of words, according to length, in the French vocabulary.

On the basis of Littre's dictionary, it has been calculated that the average length is 8.3 letters. This average is markedly higher than that of the words in a corpus which is only about 4.2 to 4.5. This is because, in the dictionary, short words appear only once, while in a corpus their frequency is high; it is so high that it largely compensates for the elongation of words due to their grammatical inflection; in the dictionary, verbs are in the infinitive, i.e. in one of the shortest forms.

A system of abbreviation which acts on words from eight letters up inclusive, covers more than half the French vocabulary. Here are the rounded-off totals, by length of word, starting from words of 7 letters:

7 letters	10,000 words
8 "	11,000 "
..... average = 8.3	
9 "	11,000 "
10 "	9,000 "
11 "	7,000 "
12 "	5,000 "
13 "	3,000 "
14 "	2,000 "
15 "	1,000 "
16 "	600 "
17 "	300 "
18 "	160 "
19 "	100 "
20 "	50 "
21 "	15 "
22 "	13 "
23 "	2 "
24 "	3 "
26 "	1 "

It is convenient to repeat here that this list does not take account of conjugated verbs, nor of plurals: the s of the plural alone causes thousands of words to pass into the class one higher; from seven to eight

letters, from eight to nine, etc. All verbs of the type finir possess forms like finissent or finissait, which count four letters more, or like finissaient, which counts six more. The system of abbreviation proposed has a very broad utility.

However, inflections do not cause more than just a few ambiguities: for example, finissait, word of 9 letters with an imperfect inflection is confused with only one other word beginning with f- and having i as second letter and n as third: finassait, which is not even mentioned in a dictionary of five thousand words.

APPENDIX III

The French keyboard.: Variations according to the rank of the letter.

An alphabet of thirty-nine signs is necessary for the correct notation of French. Here is the one used by the "Centre de Besancon":

a à â b c ç d e é è ë ê f g h i ï î j k
l m n o ô p q r s t u û ü û v w x y z

But these signs are not used in all the places of a word. The "grave" accent of à and û serves only to distinguish very short homophonic words:

a, from the verb avoir, and à, preposition,

la, article or pronoun, là adverb

ou, co-ordinating conjunction, où adverb of place.

These signs do not figure in any other word.

The sign û appears only as a final letter: it becomes a useless sign if we represent the final trigram by a code and, anyhow, it is superfluous among the initial letters.

In addition to à, û and û, there are other letters which do not occur initially:

ç, è, ë, ô, û.

The sign ç is found as the initial letter of ça only.

As for ï and î, they are met with in ïambe and île, but they are not indispensable, for there is no ambiguity. One can just as well include these two signs in the "initials" keyboard, as exclude them.

Thus the "initials" keyboard becomes:

a â b c d e é ê f g h i j k l
m n o p q r s t u v w x y z

which is twenty nine signs or, if we add ï, thirty.

We must note that the "Centre de Besancon» has omitted the sign œ, which we find in œdème, œsophage. It could be advantageous to reinstate this sign.

Thus we arrive at a keyboard of thirty one signs. If there is any point - and we believe that this is effectively the case - in following the standard of thirty-two signs, we could use the thirty second as an initial letter for the indication of capitals.

Inside a word, the signs ç and è, are indispensable, but it is remarkable that ë, ô, û, no more than î do not perform the role of discriminants. Only ï is useful, without being absolutely indispensable.

Under these conditions, the keyboard for the interior of a word can be formed of:

- (1) the twenty nine signs of the "initials" keyboard
- (2) the two signs added optionally: œ and ï
- (3) the signs: ç and è

The total being 33, we must discuss the relative usefulness of ï and of œ.

But it would be interesting to examine the usage of letters and to determine the necessary keyboard for each rank. For example, with letter II, the spread is so small that we have been led to prefer letter III, for the combination of letters II and III gives nearly as good results as the whole initial trigram.

Research will be undertaken to determine what are for each rank, the signs which never appear or which do so only in the words which can be left out of the dictionary.

We are thus led to the apparently new idea of alphabets varying according to the rank. Towards this end, the restrictions of use pertaining to final letters, penultimates and initial letters are already being studied. The practical application of these studies will require some new ideas in perforating machines: a switching mechanism will have to be interposed between the operator's initiating action, and the perforation of the letter. Moreover, anyway the proposed system of abbreviation implies a programme preparatory to perforation.

The principle of multiple alphabets, variable according to the rank, will be applied also to the case of words of three to seven letters.

The switching mechanism, even if it is an expensive one, represents a profitable investment with regard to the economy, corresponding to the suppression of half of the perforations.

Comparative Cost Evaluation

The Besancon keyboard having 39 signs considered as equiprobable, a code of five holes is insufficient, and it is necessary to have six holes. The other positions are used for punctuation, so that the Besancon keyboard has reached a total of 59 signs.

With these conditions, each letter costs six units and a word of eight letters costs $6 \times 8 = 48$. A word of fifteen letters, for example, costs 90 units.

If the procedures described above are used, we get, for a word of eight letters:

1/32	1/36	1/16	1/36	1/512
(LETTER I)	(LETTER III)	n	(LETTER IV)	(trigram suffix)
i.e. $5 + 6 + 4 + 6 + 9 = 30$ units.				

Sacrificing the signs ë, ô, û and æ for the middle of the word, economises by two units, bringing the total down to 28.

The economy thus is represented by: $48 - 28 = 20$ units
 For ten letters, it will be: $60 - 28 = 32$ units i.e. better than 50%.

If the letters IV or V are brought in, the cost changes to 33 units, which, even for a word of eight letters, shows an advantage. And, as has been said, the length is constant.