Notebook No. 2

A PROPOSED REVISION

IN THE

STRUCTURE OF LOGLAN WORDS

Incorporating the Results of Taste Tests 4 & 5

CONTENTS

The Loglan Institute, Inc.
2261 Soledad Rancho Road
San Diego, California 92109
U.S.A

FOREWORD

This Notebook is intended to allow my fellow loglanists to assess the new morphological system in its entirety and to allow consensuses to develop as to how the details should perhaps be changed before adoption by The Institute. Being a notebook, it may be updated from time to time as errors are corrected, proofs found, algorithms perfected, primitives remade, affixes reassigned, or recommended pronunciation patterns changed. So-updated, it can then become the workbook of those who undertake the extensive revision and expansion of the next edition of our dictionaries.

Many have contributed to this four-year project. Thanks are due Anthony S. Lovatt for his early insight into how the phonotactics of the 1975 language had been over-designed (TL1:183-4); to John Parks-Clifford for his analysis of Lovatt's proposal in that same issue (TL1:185-7), for his early statement of the case for H (TL1:327-9), and for his later formulation of the $(CCV)^n$ strategy of affix-assignment (TL3:273-6); to Charles J. Barton for his study of the comparative phonology of H (TL2:203-5); to Scott Layson for his clarification of the measurement-word problem (TL3:70-5); and to Jeffrey R. Brown for arguing the case for "long primitives" (TL3:120-2). Robert A. McIvor has my personal gratitude for his unstinted labor in helping me prepare, on his computer, the lengthy stimulus materials for all of our "taste tests" and for doing the first analyses of TT1-3 (TL5:111-24). My own papers in TL3:23-46, 196-200, 319-20, and TL4:5-15 complete, I think, the list of major background reading on what came to be known as GMR ("The Great Morphological Revision").

But had there not been a community of loglanists available to me as experimental subjects throughout the long period of my GMR research the system simply could not have been properly engineered...in fact, it would not even have emerged. Formal studies, while often ground-breaking and too often apparently conclusive, are simply not adequate for confronting the bristling domain of morphological fact. If there is any truth in these pages about word-goodness and -intelligibility it comes from the patient responses of those loglanists to my seemingly endless barrage of questions. Finally, I wish especially to thank Anita & John Lees, Jannaruth & Robert Jenner, and Julia & Edward Prentice for their massive response to TT4. About half the data in its tables--and so, about half of what we know about the consonant-joint--comes from these six willing people.

J. C. B.
San Diego
27 August 1982

The introductory text is divided into 15 major sections as follows:

A NEW MORPHOLOGICAL SYSTEM

by

James Cooke Brown

1. A Preview of the New Affix System: Some readers will want to know right at the top what the new affix system, especially, looks like. So let me satisfy that curiosity by saying (1) the D-Set with its CVVs and CVCs, its reduced number of "unnatural" CCVs, and its occasional local polymorphism won hands down; so D was the set chosen to be tuned; (2) the hyphenation system chosen was Hyphen R with 4-letter affixes in non-final positions; (3) final coverage by the tuned D-set was 95% vs. the 97% coverage by undecipherable affixes found in L4; and (4) the average "tastiness" of the reduced complexes remade with the tuned D-Set was 52 for 2-termers, which is the score earned by words like **dundru** (a CVC + an unnatural CCV), and 78 for 3-termers, which is the score of words like **durnortoi.** I think you will agree that these are remarkably pleasant words to be at the balancing point of a distribution. Also, the variance is low. No awkward spread remained between the best and worst words once the affixes were tuned.

2. How to Use This Notebook: You may either read this introductory text straightaway, and then settle down to a study of the listings. Or you may wish to examine the listings now, coming back to the relevant text when questions arise. Thus, you might wish to start with the affixes themselves. If you do, start with the Power Listing. This tells you more about the affixes and how they got that way than any other listing. You may also use the Affix Assignment Tables whenever you wish to look one up or study competition between them. To find out what a new-looking primitive used to be, go to the New Primitives list. To find out what old primitives were remade, and why, and what they have since become, go to the Remade Primitives listing. To see how the affixes work, go to the listing of Remade Complexes at the end of the Notebook. Questions generated by the listings will bring you back to this Introduction. You can re-enter it anywhere. All its integrally-numbered sections have been written to stand alone.

3. GMR: Its History and Purposes: The "Great Morphological Revision"--later, "Revolution", as the plot thickened--was first undertaken in the Autumn of 1978 when it became apparent that the original system for making complex predicates was not working. CPXs were not being used with any appreciable frequency in conversation, either by me or my apprentices, and when a new one was used, it stopped the conversation absolutely while one or the other of us buried his nose in the dictionary we had always to keep handy for such adventuring. When CPXs were used in correspondence, looking them up consumed more than half the time at both ends. Any systematic effort to learn them was psychologically unrewarding. This contrasts strikingly with the primitives, which go very quickly into memory, and through efforts that are easily, even joyfully sustained. Finally, when the few indispensable CPXs, like **sadja** (now **saa'dja**), were learned, they were apparently learned as quasi-primitives. The difficulty seemed to be that (i) the L4 CPXs were not uniquely decipherable; and (ii) the reduced 2-termers, like **sadja**, were not recognizable as CPXs, so not even decipherment-by-guesswork could reliably begin. 70% of the L4 CPXs were of this cryptic kind. What we needed, surely, was an affix-system that would not only produce CPXs that were uniquely decipherable, but one that would make every CPX carry its jointed meaning on its surface, like a badge.

Some other morphological tasks were soon included. As early as 1977 it had been decided (a) to include **h** in the language and (b) to "unpack" the primitive "packs", i.e., sets like the **kanta/kante/.../kanto**-set, whose members differed in only one minor sound. Also, we had had Anthony Lovatt's proposal before us since 1977 that the 1975 restrictions

on both initial and medial, but especially the medial, consonant-pairs were "over-designed": more constrained than they needed to be for intelligibility. This added (c) "Lovatt-loosening" to the work to be done. And then, in 1979, there was Jeffrey Brown's telling observation that Loglan's 5-letter primitive forms were not only Procrustean, but unacceptably so if local concepts represented by long words were ever to be gotten into the language. So, the task of (d) adding a commodious but resolvable set of "long primitive forms" to the language--for borrowing long words like 'asparagus' but also short ones like 'igloo'--also became part of GMR. Then, very recently (November 1981), the penultimate facet was carved onto the project. I began to (e) provide for the regular construction and pronunciation of "acronymic compounds", like chemical formulas. Finally, a spinoff from my solution to the consonant-intelligibility problem has now become another and, I sincerely hope, the final facet of the GMR project. This is (f) providing Loglan, which always had some consonant-clustering and may now have more (now that CPXs will be in more frequent use), with an option of "consonant buffering" suited to the needs of native speakers of languages (like Chinese, Japanese, and Italian) which tend to not to have consonant-clusters. The solution to this problem may be an unplanned benefit of the intelligibility studies, perhaps, but buffering is now a definite part of the new morphological system.

That, then, is GMR: the design of a set of decipherable affixes for the remaking of complexes that will yield as high a coverage by short affixes of the words in the present dictionary as possible; the remaking of some primitives in the interests of H-inclusion, others for primitive-unpacking, and still others for the tuning of the affix-set for maximum coverage and word-handsomeness; the design of a set of restrictions on consonant-clustering that will admit the largest possible number of intelligible combinations to the language; the installation of an optional feature for buffering consonant-clusters so that they may be sidestepped altogether when desired; the design of a set of borrowing-forms that will be as commodious as possible without encroaching on the word-space of the regular words of the language; and the design of a system of pronounceable and resolvable acronyms that will satisfy all the functions of acronyms in contemporary science.

4. A View from the Outside (with a Designer's Footnote): Let us imagine that the new morphological system presented in this Notebook is adopted. What would the language look and sound like in a few years' time? Would it then seem orders of magnitude more complex? To have lost its "stark simplicity" altogether? Let's look into this question. On the next few pages I'm going to describe the new morphology as if to newcomers...a description that might appear, for example, in the next edition of L1. Let us start with an addendum to Sec. 2.4, The Five Vowels. At the end of that section, I would add this caveat:

> There is a sixth vowel sound in Loglan, but it is not spelled with a vowel letter. So let's consider it in the next section on the consonants. The sixth sound is in fact a very short, neutral vowel that German linguists call "schwa". It occurs quite frequently in English and is the value of the 'a' in 'sofa', for example, or of the 'e' in unstressed 'the'. Surprisingly, in Loglan schwa is an allophone of a consonant, not a vowel.

Then, in the next section (which will now, of course, be devoted to The Seventeen Consonants, now that there is **h**), I would add this brief explanation of the role of **r**, one value of which is schwa. These new paragraphs might well replace the last paragraph in Sec. 2.5, now unnecessary:

> One of the four vocalic consonants, namely **r**, has a special role to play at the "joints" in certain kinds of words. At these joints **r** is used to separate two syllables which, for some good reason, should not come together. Sometimes that reason is that the syllables have distinct meanings which are to be joined together in a single notion but nevertheless kept visually and audibly distinct. For example, the Loglan word **mekrkiu** means 'eye-doctor'. Clearly, the two **k**'s should not come together or that jointed meaning would be lost. So the inserted **r**, taking on one of its two vocalic values, manages to preserve the two main syllables. It introduces a short, always unstressed burst of neutral sound that keeps the two **k**'s apart but cannot be

mistaken for an ordinary vowel. (As English-speakers, we would want to spell this sound with 'er', and transcribe **me'krkiu** into mock-English as 'mecker-cue', accent on the 'meck'.) In this position **r** is rather like a hyphen, but a hyphen which appeals not only to the eye but to the ear.

Now unlike the five true vowels sounds of Loglan--which are clear, mouth-stretching sounds as in Spanish--a good deal of latitude is allowed in pronouncing this hyphenating vowel. Even so, it is always spelled 'r' no matter how it is pronounced. Speakers of most American dialects of English will usually prefer their own "vocalic **r**" for the audible hyphen, pronouncing it just as they do the '-er' in 'fatherhood', 'maternity' and 'bittersweet'. Native speakers of other English dialects, and of all those many languages in which a vocalic **r** does not occur, will no doubt use their own familiar schwa: that practically universal unstressed vowel whose main role in human languages seems to be easing the burden of consonant-clustering. And that, in fact, is exactly how the suffix 'er' is pronounced in nearly all dialects of British English and in German.

Summing up, the consonant **r** has three allophones in Loglan. First, it has its usual consonantal value when initial ('red'), or when between a consonant and a vowel (as in 'tree'), or between two vowels ('era'). Second, whenever the **r**-sound is jammed between two consonants (as in **mekrkiu** or 'eternal'), or is initial and followed by a consonant (as in English 'irk' and 'Earl'), it may take on either of its two vocalic values: schwa, which is the 'a' in 'sofa' or the 'e' in unstressed 'the', or the vocalic **r** that is so common in American English and in some few other languages. So, as I trust the reader is a speaker of at least some dialect of English, the letter 'r' between consonants will look odd to your eye, but it will not sound odd to your ear. Your ear and tongue already know it. All you need to do is teach your eye to see the sound you know as 'er' in 'r', and your hand to drop that 'e'.

Now, with this small adjustment to the phonology, let's move on to the morphology. Secs. 2.8-9 on little words are largely ok as written. But Sec. 2.10 on Predicate Words will have to be entirely redone. Here's a first draft:

2.10 Predicate Words

Predicate words form the bulk of the vocabulary of any language and range in length from short, frequently used words, like English 'egg', 'run' and 'boy', to very long, seldom-used predicates like 'antidisestablishmentarianism'. In Loglan, too, one can string forms with separate meanings together and so express extremely complex ideas in single words. Words like 'infix' and 'understand' illustrate this process at its earliest level in English. 'Understandable' and 'infixive' go one step further; and so on.

But let us first consider the forms of those Loglan words that express the kinds of concepts usually written indivisibly, such as 'egg', 'run' and 'boy'. One would expect such words to be short in Loglan, and they are; but not as short as the shortest predicate words in English. For Loglan is a much smaller language phonetically than English; and so it has fewer short words. Besides, nearly all the monosyllabic words of Loglan are heavily-worked structure words like English 'of' and 'the'. Words with meanings like these occur more frequently than even the most common predicate words, and so deserve to be the shortest words of any language.

The indivisible predicates of Loglan are of two kinds. They are either words borrowed directly from some single natural language, like **iglu** is borrowed from Eskimo in both English and Loglan, or they are composite predicates built up of overlapping sound-sequences taken from as many of the eight target languages as possible, thus making them as recognizable as possible to the world. All the predicate words you have seen so far (except of course **iglu** and **mekrkiu**) are of this internationally-derived composite kind. Thus **junti, mrenu** and **botci** are all composite predicates. It is no accident that they are all five letters long. All the simple composite predicates of Loglan are constructed in one or another of just two five-letter forms:

CV'CCV or CCV'CV

an arrangement that can be summarized as

CV'C/CCV' + CV

The internationally-derived predicates of Loglan are, as you would expect, the semantic building blocks of the language: its primitive notions. So every notion which is universal enough in human experience--as igloos are not--to have found expression in a simple word in nearly every human language is expressed by a composite predicate in Loglan. Thus knives, boys, eyes and healers are universal in human experience; so each has its own simple predicate in Loglan: **najda, botci, menki** and **kicmu.** (You can probably sense the international origins of all these words.) But perhaps eye-doctors and knife-boys are not. So these more complex notions are, as befits a logical language, expressed in Loglan by more complex words. Thus words like **mekrkiu** and **najboi** ("nazh-boy") arise. We will see how to build such words out of parts of primitive predicates in a moment.

But neither are igloos, llamas, asparagus, chlorine or australopithecines. So in addition to its apparatus for building complex predicates out of simple ones, Loglan also has an apparatus for borrowing predicate words directly from single linguistic sources...in particular, from the languages of the people who do have igloos, llamas, asparagus, chlorine, or australopithecines. Most importantly, perhaps, the international vocabulary of science, in all its fullness, must be allowed to enter the language without significant distortion. Some distortion is inevitable. Every language distorts what it borrows. But let us see how distortion may be minimized in a neutral language.

Some words we wish to borrow just happen to fit one of the two word-forms already assigned to composite predicates. The Swahili words **simba** for lions and **dumbo** for elephants clearly do; the virtually untranslatable Hindi word **karma** does; and 'chlorine' transforms with very little loss into **clori.** (**Clori** is pronounced 'shlaw-ree', of course. It is the spelling we will usually want to preserve in importing scientific words.) Similarly, the international word 'telephone/-fono' is neatly compacted as Loglan **telfo.** And what about **telvi** for 'television' and **futbo** for the international game that most people who play it spell 'futbol'? It would be foolish to deny these handsome words entrance into our neutral language on the grounds that they "look like" Loglan composite predicates. No problems are created by the fact that they are not.

But Loglan has a second and much more flexible set of word-forms designed to accommodate more difficult borrowings. Like those of all Loglan word-classes, the forms of these borrowed words, too, are distinctive. The shortest of these "borrowing forms" are 4 letters long (**iglu**) and they go up through 7, 10 and beyond by increments of 3. In the language of mathematics, they are all 1 mod 3 letters long. Like all regular Loglan words (i.e., non-names), they end in vowels. Also, like the composite primitives but unlike structure words of whatever length (which can have no adjacent consonants), all these borrowing forms must have at least one consonant-pair within their first four letters. That is a quite general sign, in fact, of a predicate word. Also like all predicates, borrowed predicates are all stressed on their penultimate syllables, that is, on the syllable next to the last. This is the commonest stress-pattern in Loglan. Predicates share it with many other sorts of words.

Examples of borrowed predicates which have been slightly altered to fit these 1 mod 3 forms are **spai** ('spy', this one is phonemically identical to its original), **e'lki** ('elk'), either **o'ksi** or **oksi'gne** for 'oxygen', **engli'ca** (for 'English'; but please put the stress in the right place), **arkni'da** or **rakni'da** (for 'arachnid'), **asparagu'sa** ('asparagus'), **krustei'cia** ('crustacean'), **australopi'tku** ('Australopithecus'), and **anglosakso'nia** for 'Anglo-Saxon'. (I am marking stress here so that you will be able to pronounce these new words properly the first time you see them. Later, you will not need this help.) All these words are predicates despite the fact that some of them are capitalized in English and so seem to be names; and all have precise local definitions or established usages in science. And remarkably enough, despite what appears to be their uncontrolled variety, all but the **spai**-form words are described by a single formula:

$$(CC/VCC/XVCC + (X)^{n}) + V'C(C)^{m} + V/vv$$

In this formula, n and m give the number of instances of the marked elements. Their values are chosen to produce a length of 1 mod 3, and either or both may be zero. An X may be either a consonant or a vowel; so the middle portions of long words may be quite freely contrived. vv is either a diphthong (like **ei** or **ao**) or a pair of vowels capable of being pronounced monosyllabically (like **ia** or **ui**). The components in parentheses are optional except that if m = 0 (thus leaving a single consonant after the stressed vowel), then the first term (with its guaranteed consonant-pair) must be chosen. **spai**-type words have the simple formula CCvv, and so are always monosyllabic (like **spai** itself). It is an important restriction on both these formulas that any final vowel-pair be monosyllabic. For look what happens when one is not. Take ***spea**, for example, in which **ea** spans two syllables: /SPEa/. Then the phrase **to spea**, or two of whatever "speas" are, would come out /toSPEa/ and so be indistinguishable from the complex predicate **tospe'a** (made up of **tos** + **pe'a**). This is intolerable in Loglan. Among its other properties, Loglan is to be machine-intelligible. So no "I scream/ice-cream"-type problems are to be left lurking in the language. Thus, words of ***spe'a**-form are not allowed. There are some more restraints on word-borrowing in Chapter 6.

Let us now consider the forms of complex predicates: the ones that are to be derived within the language from its own primitive predicates. We have seen three instances of these jointed words: **me'krkiu** with its audible hyphen, **na'jboi** which needs no hyphen, and the one we have just stumbled on by considering how 1 mod 3 forms should **not** be built, namely **tospe'a.** (I will continue to mark the stress in these words until sensing its location becomes automatic for you.) It is always the penultimate true vowel or monosyllabic vowel-pair in a predicate that receives the stress. Thus **oi** (pronounced 'oy') and **iu** (pronounced 'you') are monosyllabic; **ea**, we have seen, is not. So the stress slips to the right in /toSPEa/ and stays firmly left in /MEKrkiu/ and /NAJboi/. Note that hyphen **r** is not even counted in figuring stress. In fact, hyphen **r** isn't counted at all. It might as well not be there as far as the structure of these words is concerned.

Now you will find neither **na'jboi** nor **tospe'a** in any dictionary. But that doesn't matter. We would know immediately what any speaker who used one of these words probably meant. **naj** is an affix (a combining form) that can come only from **najda** = 'knife'. **boi** is an affix that can come only from **botci** = 'boy'; **tos** comes from **tosku** = 'skull' and **pe'a** from **penta** which means 'a point' or 'something pointed'. Now you might not know exactly what the speaker has in mind by speaking of "knife-boys" and "skull-points", but you are now in a very good position to find out. Thus, complex predicates in Loglan, as in German, are semantically transparent. No German child needs to ask whether 'Handschuh' means those items of apparel that we call gloves. Of course it does. What else could a hand-shoe be? So it is in Loglan. Once you know the affixes which have been assigned to all the primitive predicates of the language--about 800 of them are currently assigned--you will be able to decipher at first sight every new complex word you will ever encounter.

Take the 12-letter word **rojmadse'smao.** Such words will soon break up before your eyes...or in your ears, for that matter. Perhaps either your eye or ear has already told you that this one can only be the 4-term complex **roj** + **mad** + **ses** + **mao.** You have already guessed that all these glued-together words are 0 mod 3...not counting hyphens, of course. So you can chop their syllables off in 3-letter segments. The first such segment, **roj**, as you are now learning, comes from **rodja**, 'grow'; **mad** comes from **madzo**, which means '(to) make' or 'a maker'; **ses** comes from **sensi**, which means 'science'; and **mao** (which rhymes with 'cow') is yet another combining form of **madzo.** (Some primitives have two or more affixes, each useful in different contexts.) Now what on Earth could the speaker or writer mean by a "grow-make-science-maker"? Of course. An agronomist! What else could it mean? And **se's-mao**, of course, must then be 'scientist'. Nearly all the metaphors behind Loglan complex predicates are of this transparent kind. So dictionaries are, in fact, of little use to Germans and loglanists once they "know their affixes".

By the way, although a primitive may have several affixes, every Loglan affix is assigned to just one primitive. So once you have learned that **mao** (still 'cow') is a

contraction of **madzo**, you are done with **mao**. It will always mean **madzo**. So there is no such thing as an "ambiguous affix" in Loglan. Does 'inflammable' mean 'capable of flaming in'? Or 'not capable of flaming'? Does 'infix' mean 'fixed in'? Or 'not fixed'? It means 'not flexible' in 'inflexible'! Such questions cannot even arise in Loglan.

We have uncovered three kinds of affixes: the CVC-forms of which **mad, ses, mek, tos, naj** and **roj** are instances; the monosyllabic Cvv-forms of which **mao, kiu** and **boi** are instances; and the disyllabic CVV-forms of which **pe'a** is still our only instance. There is a fourth contraction, and then four longer forms we must consider. Take the predicate **mrenu**. The obvious choice of a 3-letter contraction for this word is **mre**. Another and very frequently used CCV-form is **cli** from **clika**, which means 'like'. So **mre'cli** must mean 'man-like' or 'manly'. Many American Indian languages have a word that means 'man-woman'. We can convey this local metaphor very neatly in Loglan with **mre'fua**, in which **fua** comes from the word for 'woman', which is **fumna**. (Pronounce **fua** like the "fwa" of French 'foi', not as "foo-ah".) With the CCV-form we complete the list of short affixes, or contractions.

Now nearly all Loglan primitive predicates have contractions. But some do not; and some that do, may not have the right kind of contraction for some spot in a complex. When this happens, an "unreduced", or long affix, must be used. For example, **mubre** ('wood' or 'wooden') has no contraction. So 'lumber', or 'building-wood', is **bacmu'bre** in Loglan; and in this word the short affix, **bac**, comes from the primitive **balci**, '(to) build'. (**bal** is a contraction of something else, namely **balpi**, 'balance'.) Similarly, **banko** ('bank') has no contraction. A banker is a "bank-doer" in Loglan. So the word is **ba'nkrdru**. We needed that hyphen to "glue" the word together. Note that the final **-o** of **banko** has simply been replaced by the audible hyphen, a much shorter sound. **dru**, of course, comes--rather irregularly, it turns out--from the word for 'do', which is **durzo**. It is because **durzo** is such a frequent and mobile component of complex predicates that it has been given this irregularly-derived but powerful CCV affix. CCV is the only affix-form that will literally "go anywhere".

We now have all eight of the forms that can be strung together to make the complex predicates of Loglan. They are CCV, CVC, CVV and Cvv, which are the four contractions, and CVCCr, CCVCr and the two 5-letter primitive forms themselves, CV'CCV and CCV'CV, which are the four long forms. Clearly, the 5-letter forms may only be final; and the 4-letter forms (with their accompanying **r**'s) may only be non-final. The CVC-contraction, too, is never final.

The monosyllabic Cvv-form is a little freer. It may go anywhere in a complex except at the head of 3-term or longer complexes. There it would come unstuck. For example, ***maomre'fua** wouldn't be a good word because the first syllable would sound like a separate word to the listener. Inevitably he or she would hear the intended word as a pair of words, **Mao mre'fua**, which happens to mean 'Mu is a man-woman'. This is the "I scream/icecream" problem again. So Cvv's are not permitted in such positions. However, a Cvv works perfectly well at the head of a 2-term complex, as in **boi'mre**. In that word, **boi** can't come unstuck. What would be left if you took **boi-** away is **-mre**; and ***mre** is not a word in Loglan.

The disyllabic CVV-forms like **pe'a** are the most restricted of the four types of contraction. They may used as final terms in 2-term complexes provided the first term is either a CCV or a CVC (**mrepe'a** or the **tospe'a** we have already seen); or they may either be final or penultimate in longer complexes (**mrefuape'a** or **mrepea'cli**) but never earlier than that. Also, two CVVs may not be adjacent in a word.

I have already mentioned that CCV is the only combining form that is free to go anywhere. So a word may be composed entirely of CCV-type contractions. **Mrecli'dru** is such a word. (Can you decipher it?) There is thus an extraordinary variety of complex predicates in Loglan, and the formula describing all of them is naturally quite complex. It is given in Chapter 6, where complex-word making is discussed in more detail.

Summing up, there are the composite primitives of Loglan, which are either of **mre'nu-** or **fu'mna**-form. Some borrowed words like **clo'ri** and **si'mba** also have these 5-letter forms. Then there are the 1 mod 3 borrowings ranging from **spai** and **i'glu** at the short end to **australopi'tku** at the other. And finally there are the complex

predicates of Loglan. These are 0 mod 3 (not counting any hyphens) whenever they are in fully reduced form. They range from 6-letter words like **na'jboi** and **mre'cli** to words of any length whatever depending on what terms have been left unreduced. Thus the form of 'agronomist' in which all terms are left unreduced is the 20-letter monster **rodjrmadzrsensrma'dzo.** It is difficult to imagine circumstances in which anyone would use this word. On the other hand, someone might. So it is good to know how to break it apart. If you think of the **r**'s as real hyphens, it breaks apart very easily: **rodj-madz-sens-madzo.**

Before leaving the predicate forms of Loglan, it should be pointed out that no one is ever obliged to use reduced complexes. Words composed entirely of contractions are often neat and quick, and with knowing listeners in conditions of low noise, they are usually the best choice. But if the audience--say a learner--is unfamiliar with the concept, or if there is a good deal of noise, then a longer version of the same word is likely to be a better choice. Consider the following sequence:

sa'npa dja'no
sanprdja'no
sa'nprdja
saa'dja

All these expressions mean the same thing. The first is the original metaphor, "sign-know", say at the moment of its introduction into the language by an innovative speaker. (That happened twenty years ago). The second expression is a single word composed of two long affixes with just one stress, and is definitely shorter. Later, with increasing use of the concept, the third expression might become most common. It has now become /SANprdja/, an even swifter word. Finally, as the concept in question (which in Loglan means 'understand', as in '(to) understand the meaning of some sign (not a person)') became very widely used (as in fact it did, about five years ago), its final form would be reached: /saADja/. And that is as short as this predicate is going to get, no matter how frequently it is used.

But this same developmental sequence is available to be used by contemporary speakers for quite different purposes. In different rhetorical circumstances, say, or in noisy conditions, a speaker might switch from **saa'dja** to **sa'nprdja,** or back to **sanprdja'no,** or even go back to the metaphor **sa'npa dja'no** itself. To give three examples, he or she might be (i) giving a formal lecture on "understanding" to university students, (ii) teaching the language to a pupil who was just learning this concept, or (iii) shouting in a storm. The point is that, in Loglan, the speaker is free to do any of these things. All these distinct versions of the "same word" coexist for him in the language side by side, so to speak. Anyone who understands any of them will eventually understand them all.

Thus all Loglan complex predicates are polymorphic, a feature of the language perhaps especially well-suited to second-language learning by adults, or to a language which is meant to grow.

Well; what is your opinion? Is the morphology of the new language going to seem "orders of magnitude more complicated" to the incoming loglanist circa 1985? The trunk of the morphological tree is still the same: the primitive predicates. And these are by far the most numerous words encountered by a newcomer. As for the farther and more far-reaching branches of the word-tree, is it really going to bother anyone to occasionally encounter words like **iglu** in Loglan? Or **australopitku**? Does it bother you in English? No sophisticated conversation, no trip to the human zoo, is possible without them. And given the knowledge that in these encounters you are standing in the morphological doorway through which the whole of science might one day enter Loglan, I rather imagine that you will come to welcome these borrowings. They are clear signs of the worldliness and capaciousness of our language. And what about the complex predicates? Are these decipherable ones--with their little gleaming beads of constant meaning, strung together in the plainest of ways, unstringable at a moment's notice--**really** more complicated than those baffling ones we used to live with (and which nobody liked to use)? Which no matter how you pushed and pulled at them simply would not come apart? Never reliably, anyway. You could never be sure that you were carving **those** pieces of reality at the

joints.

I think not. I think the complex predicates of 1975 Loglan were not very simple at all. Just because we didn't talk about them didn't make them simple. If I **had** written about them--tried to explain how all those little bits and pieces went together, and why one piece meant one thing one time and another thing another--it would have taken a notebook twice this size. And then made no sense. Nope; we've now got a simpler language, Rorpern. Just because the **engineering** was intricate doesn't make the object that it led to intricate. The hull of a boat is simple. But the equations that describe the balancing act between its many necessary virtues are most wonderfully complex. Be forewarned. The design-studies behind these simple affixes are similarly complex.

5. A Glimpse of "The Word-Maker's Manual": This is only a glimpse because there isn't room for more. Besides, Bob McIvor's final testing of the resolution algorithm, which will be the foundation of that manual, has had to wait for the completion of the rest of GMR. His algorithm will be published separately, probably in TL. Nevertheless, I can summarize what I have learned from making the three sorts of new words:

5.1 Making Complex Predicates: There is only one problematic sequence that I know of, and it is a generalization of the old ***Tosmabru** Case. Words of form $(CVC)^n$ + CV'ccV, n greater than 0, in which cc represents an "active initial", i.e., one of the 36 CCs used in making CCV-affixes; see CCV Assignment Table. Every word of this form must be checked to make sure that not all its C/C joints are "bridged" by active initials. If all are, it will break up as a CV # $(ccV)^{n-1}$ + ccV'ccV, e.g., as **to sma'bru**, and is disallowed. If any joint is not bridged, it's ok. So it is often possible and always sufficient to replace one of the CVCs in such a form with an allomorph that will destroy one of its bridges. Suppose we wanted to make **tosku matma setci** ('skull-mother-set', no doubt a kind of secret society common in the primitive fastnesses of Loglandia) and had written ***tosmat-setci** as a first try. Noticing that SM, TS and TC are all bridges, causing the sequence to break into **to smatsetci** ('two smoke-error-eaters'!), we might replace **mat** with **mam**, also 'mother', and write **tosmamsetci**, which works fine. (M/S in ***smamsetci** is not bridged.) Of course, a more obvious solution (since **sei** exists) would have been to get rid of the CV'ccV-term, with its fragile center, and to have offered the Loglandic anthropologists **tosmatsei** (/tosMATsei/) in the first place.

It might be thought that CPXs such **tosma'tsei, tosmao'sei** and **tosmaa'sei**--in short, any 0 mod 3 form with initial CVCC- and final -Cvv, with the first CC an active initial--would break up into phrases like **to sma'tsei**, which certainly look like 1 mod 3 type borrowed predicates preceded by CV operators. Indeed they would break up in exactly such ways if such forms were permitted in the borrowing lineage. But all 1 mod 3 forms which might abridge the right of CPXs to occupy their own word-space have been quite deliberately excluded from the borrowings; see Sec. B below.

One quite surprising outcome of my resolution-work with CPXs is that ***sea'dja** won't work but **saa'dja** will. CVV'+CCV words won't work **unless** the two adjacent Vs are identical because they break up immediately into CV # VCCV phrases if they're not. Thus **se adja**, or 'seven "adjas"' (whatever they are), is how /seADja/ will be understood no matter what the intent of the word-maker. But the old La Ailin Rule requiring a mitigating pause--normally a glottal stop--between the **a**'s when the name operator precedes an **a**-initial name, thus /la.aiLIN/, is most reasonably extended to protect the new VCCV-form words from the same kind of adventitious vowel-doubling. Thus, it is most natural when saying **sa adja** ('almost all adjas') to stop glottally between the two **a**'s, and say /sa.ADja/. So **saa'dja** spoken without such a pause comes out laughing. It must be a complex. So I've simply proscribed CVV'CCV-forms without doubled V's, and allowed the ones with doubles through. Cvv'CCV-forms like **tue'dji** work just fine, of course. This is one of the reasons why the Cvv's are so much more valuable in the affix-set than the CVVs.

5.2 Making Borrowed Predicates: As explained in TL3/4, pp.319-20, the 1 mod 3 group of forms is a residual class. That is, it is to be contrived in such a way as to include as a legitimate borrowing any V/vv-final, 1 mod 3 sequence with an early CC that is not also (1) a possible composite plus a preceding CV, or (2) a possible complex either (2a) plus a preceding V or (2b) less its initial CV. ***tobrudi** illustrates the first exclusion, CV +

ccV'CV, ***osmacli** the second, V + CPX, and ***smatsei**, which we may write CPX - CV, the third. (That is, ***smatsei**, as we saw in Sec. 5.1, above, is the CPX **tosmatsei** less its initial CV.) But now please note that **oksigne**, which I listed as an alternative to **oksi** for Oxygen in Sec. 4--and which certainly has the external form of ***osmacli**--is an acceptable borrowing. It will not break up as **o *ksigne** because ***ksigne** is not an acceptable CPX. So, clearly, if we wish to make our residual class as large as possible, we must take into account the "un-bridged" consonant-pairs within a borrowing that might save it from either breaking up, or coalescing, into something else.

For borrowings of length 7, all the residual forms were once worked out, but without taking into account the freeing effect of "non-bridges". Thus, **oksigne** and **raknida** were then excluded. Here is the new set, almost certainly incomplete. In the new formulas 'C-C' stands for a non-bridge, i.e., a CC that is not one of the 36 active initials. A 'CC' may or may not be an active initial; a 'cc' is an active initial; a 'VV' may be a vv, and a 'vv', of course must be:

1) VccV'C-CV 2) (CV/VV) + C-CV'CV 3) (VC-C/ccV) + V'C + (CV/vv)
4) (ccV/XCC) + XV'CV 5) (XVV'/XV'C/V'CC) + CC + (CV/vv)

Notice that ***sma'tsei** and ***osma'cli** are not among these strings and **oksi'gne** and **rakni'da** are. **Oksigne** is String #3, as so would be **oksi'gia** and **oksi'ste. Rakni'da** is an instance of String #2, as is **iodnina.** Excluding them presupposes that primitives like ***knida** and ***dnina** could exist, and so unnecessarily reduces the borrowing space. Similarly, **oski'gne** (with an **sk**) is String #1, in which the first CC is bridged but the second isn't. A late non-bridge is just as good as an early one to protect a borrowing. Thus, **oski'gne** is a good borrowing precisely because ***skigne** is not a good complex. I would welcome the complete residual set for the 7's, plus a proof that it is complete, from any loglanist willing to take on this tricky problem.

For the 10+'s, it is only necessary that words made to the formula given on p.5 be checked for the 4 following conditions: (1) Is the "ending", i.e., the post-stress sequence starting with the first C after the stressed vowel, of length 3 or 4? In particular, is it any of the forms CCV, Cvv, CCCV or CCvv? (2) Does it start with VC? If both of the first two conditions are met, then blot out the initial vowel and look at the rest of the word. (3) Is it a CPX? (4) Could it be made a CPX by appending any V-final sequence to its front? If it is or could be, then that particular borrowing is not allowed. (In a 10+ word it is usually easy to make some changes in the middle sequences that will make it allowable.) Note that if either of the first 2 tests fail, the last 2 will also; so the word is allowable.

The set of 1 mod 3 words of length 10 or greater that pass through this screen may not be a residual class in the strict sense that they include all the sequences that meet the other requirements of a predicate; but they are an immense number of words. Bob McIvor's word-checking algorithm, when it is complete, should give us criteria for a larger, if not yet actually complete, set of 1 mod 3 borrowings. However, the word-maker, when contriving a long borrowing, will often discover a sequence that will actually meet the functional test of a good borrowing even though no published formula allows it. Please communicate all such discoveries to The Institute.

5.3 Making Acronymic Compounds: The acronyms of Loglan are not only what acronyms usually are, that is strings of (usually) capital letters written together and surrounded by spaces, which exhibit in order the "important" letters, often initials, of some much longer word or phrase; they are also compound little words. They thus fall in the same morphological class as compound tense-operators or number-words. Thus **AAA** is the acronym; **AcAcA** (pronounced /aCAca/) is the acronymic compound. Note that stress is penultimate. The writing convention, of course, is to capitalize in the compound the same letters capitalized in the acronym. **-c-** is the acronymic hyphen (not **-z-**, as reported in the McG Notebook). Its function is both to incorporate in the compound, and to render more swiftly pronounceable, the vowel-singlets (usually) used to represent the vowel-letters in the acronym. Thus, /eh-eh-EH/, in which the vowel syllables are uninsulated from each other and also diphthongs, takes more time to say than /aCAca/; try it.

Any consonant-letters in an acronym are spoken as full letter-variables; these then appear as full CVV-syllables in the compound. Thus **CaiCaiCai** is CCC written out; and

/caiCAIcai/ is the pronunciation of both expressions. The listener's default assumption is that any vowel-singlets in an acronymic word are Latin capital vowel-letters in the acronym...except in chemical symbols, in which any singlet representing a sutori letter in the acronym is assumed to be in Latin lower-case. Thus **CaiIcA** is the phonemic form of **CIA**, and both are pronounced /caiICa/. (Alternatively, /caiCICa/, should it be thought we need that second hyphen.) But if the listener knows that **Acu**, which he hears as /ACu/, is a chemical symbol--the symbol for Aurum (gold)--then he knows that he must write it acronymically as **Au**. Single-letter "acronyms"--which are not acronyms at all but letter-variables--such as often occur in chemistry, e.g., **N** and **O**, are spoken aloud as full letter-variables, e.g., **Nai** and **Oma** (/OMa/). And **DNA**, of course, is either **DaiNaiA** or **DaiNaicA** with a hyphen. (But I can hear the difference between /daiNAIa/ and **Dai na ia** as /daiNAia/ without that hyphen; can't you?)

The use of **-c-** as the vowel-hyphen in acronyms has only one known constraint. If a phrase to be rendered as an acronym requires two initial vowel-letters of which the first is 'I'--suppose the acronym required were a transliteration of 'YACC' as **IACC**--then the normally-constructed compound doesn't work. For ***IcACaiCai** (/icaCAIcai/) breaks up as **Ica CaiCai**, which means 'Or CC'. In such cases, the first of the two vowels must be spoken as a letter-word in pronouncing the acronym. Thus 'YACC' in Loglan comes out **ImacACaiCai**, pronounced /imacaCAIcai/...a bit longer, for once, than /eye-eh-see-SEE/. But apart from the **i-** in **ic**V-form connectives, no other single-vowel word precedes a **c**V-syllable in grammatical Loglan. (Sheks may not, of course, grammatically follow eks.) So **OACC**, for example, works out just fine: /ocaCAIcai/, which again is considerably shorter in speech than the English /oh-eh-see-SEE/.

Remarkably enough, **Irei**-type words--**Ir** being the symbol for Iridium--are held quite harmless. It is a generalization of an old stress-rule that handles this case, namely that (now) no single LW, nor final syllable of a compound LW, may be stressed as a pauseless antecedent of any other polysyllabic word (not just a predicate). Whence /IRei/ cannot now be heard as **I rei** which, if pauseless, must be spoken either /irei/ or /iREI/. If the speaker does wish to emphasize the connective, he must pause after doing so...quite a natural thing to do. Thus **I rei groda sei** with stressed **I** comes out /I.reiGROdasei/. This is, in fact, exactly how we say it in English: '**And** (pause) r is bigger than s'. Acronyms and acronymic words are given a thorough work-out in Part 5 of the Remade Primitives listing.

6. The Series of Trial Affix-Sets: The major movement of the GMR research over the last four years is described in the way the trial affix-sets, and the strategies for making them, have changed. Seven sets have been made, not counting the tuned version of Set D. Let us call the first 3 sets (then unlabeled) Sets 1, 2 and 3. The last four were the labeled Sets A, B, C and D. A and B were the pair compared in Taste Test #1; C and D in Taste Test #5.

Sets 1 and 2 were partial sets made for the first of the GMR studies: the 1978 study of coverage and remaking-cost. Set 1 was to provide CV-form final affixes for the 85 prims that were most productive in this position; and filling the CV-table in this way required that many of those primitives be (provisionally) remade. Set 1 also supplied either a CCV affix or a pair of complementary CVCs to each of an overlapping list of prims that were most deserving in non-final positions. Set 1 preserved, therefore, the (non-)virtue of 2 mod 3. Set 2 introduced the CVV-form, and used both these and CCVs in final positions. It provided complementary CVC-pairs as well as CCVs for non-final use. Set 2 thus introduced the 0 mod 3 concept which assured that all CPXs would be recognizable as such, and was thus the parent of all future sets.

Quite apart from its resolution of the recognition problem, the coverage vs. remaking figures also favored Set 2. Set 1 would have required more than twice as many remade prims as Set 2. It was also probable that, in the end, Set 2 would also have had greater coverage. That point had not been reached, however, when the study was terminated at the 75% coverage-mark for both sets. It was clear that both would reach the 90's.

Trial Set 3 was built to explore an orthogonal strategy, namely that all short affixes be CCVs. The complex word would then have the attractively simple formula $(CCV)^n$...if all its terms were covered. But this condition was not frequently to be met. Investigation showed that a term-coverage figure in the low 80's could not be exceeded even if some 30 of the most powerful (and handsome) primitives in the language were

remade: primitives like **matma, fumna, ganta** and **takna** which yield no CCV affix with a permissible initial no matter how leniently permissibility is defined or how much reordering of phonemes is allowed. Unless those 30 "impossibles" were remade, and in most cases, radically, coverage could not exceed the middle 60's. Thus, on the grounds of coverage alone the $(CVV)^n$-strategy could be rejected as illusory. At 65% term-coverage, most complexes would not be $(CCV)^n$; and at 80% scarcely more than half would answer this description.

But the decisive consideration was the huge loss in recognizability, not to say handsomeness, in the primitives themselves if remade to get the higher coverage. Thus, in the course of filling its CCV table, $(CCV)^n$ would have forced such words as **zgaha** upon us, which was **ganta** remade to fill the **zga**-slot in that table. **Zgaha** suffers a large R-loss; and is certainly not very pleasant either to look at or say. Yet it was the best composite word for 'high' that would fill any open slot in the CCV-table by the time **ganta**'s turn on the power list came around. **Ganta** was not atypical. In fact nearly all the 30 impossibles--which are in the main, like **ganta**, handsome, high-R words--would have had to be made less recognizable in this unnatural way. Indeed, the consonant-combinations that would thus be installed in some of the most frequently-used words in the language would be precisely those that do not occur with any frequency in natural languages. That's what happens when you exhaust a table of possibilities. So the new prims would not only have been unrecognizable and, collectively, would not have covered much, they would also have been strange.

So the search for decipherability continued along the path of polymorphic affix sets. Both Sets A and B were of this kind, using all 3 types of affixes. They were also the first complete affix sets, and so the first for which coverage could actually be measured. Set A had 93% coverage, Set B 92%; both quite tolerable but not yet maximal. Set A retained all of the CCVs of Set 3, however unnatural (e.g., **sfa** from **setfa**), except those which were to be achieved by the remaking of the "impossibles". In short, A maximized the number and power of its CCV-affixes, but without bizarreness, by using CVCs and CVVs only where no CCV could be more powerfully used.

Set B in contrast, abandoned a good many but not all of the unnatural CCVs that had been created for Set 3, using only naturals or "good unnaturals" (e.g., **fra** from **farfu**). It also used complementary pairs of CVCs quite lavishly (e.g., **mat/mam**), often augmenting even these with a CVV (**maa**, in this case). For less powerful prims it used single affixes: a CVV, a frontal CVC, or some new medial one (**rel** from **trelu**).

TT1 clearly showed the B words to be more tasty than the A words but without pin-pointing the cause; see the discussion of the TT1 results in Sec. 7 below. But an AB vector had been established by the difference in their strategies. Supplemented by the more specific insights gained from TT1, that vector seemed worth extending in the construction of Set D.

But there were other conclusions that could be drawn from TT1. Bob McIvor embodied some of these in Set C: a set made entirely without CVV-form affixes but with large numbers of unnatural CCVs, and even some "bad unnaturals" (**dzo** from **madzo**). This was a retrograde step but a useful one, in that it allowed the detailed investigation of the naturalness issue in TT5. Set C also had no local polymorphism, assigning just one affix to each primitive that had any. The predicted coverage of Set C without CVV affixes was, of course, very low, although not quite so low as Set 3...about 70%, it turned out.

In making the D-Set I pursued a diametrically opposite strategy. I reduced the number of even the "good unnatural" CCVs from Set B, replacing them with CVC/CVV couples when deserved. I had no tradeoff figures then with which to calculate the costs and benefits of these replacements, and it turned out later that I went a mite too far. (So in tuning, some good unnaturals were invited back.) I also used CVVs very freely, giving the monosyllabic ones more work to do whenever possible. I also cut down significantly on the number of CVC-pairs, having learned from TT1 how valuable these forms were, and so to spend them thriftily. Finally, I had the new 4-letter "long affixes" to investigate, and these got folded into the words made with the D-Set affixes. C, of course, used the old 5-letter-plus-hyphen plan that had been used in L4.

As TT5 exhibits in abundant detail, the sharp differences between the C and D strategies of affix-assignment produced a far more sensitive experiment than could have been achieved with stimulus words more similarly contrived.

7. Taste Test #1: A Comparison of Two Strategies of Affix-Assignment: This was the "snippet test" sent out with the Supplement in November 1979 after an earlier and more ambitious effort to compare words made with A- and B-Set affixes had failed. The version of TT1 that worked was to supply each TL-subscriber with a one-inch strip, containing about 20 trial words, half made with each affix set, cut from the list of about 4000 trial complexes computer-generated by Bob McIvor; see Sec. 6 for a description of how the A and B Sets differed. Eighty-three loglanists responded, their ratings covering 1799 words, about 900 from each set.

The results of TT1 were both puzzling and provocative. I will list only the most fruitful ones here. (1) B-words, with their fewer "unnatural" CCVs and more numerous CVVs, were definitely preferred over words made with the older A-Set (P about .0002). This suggested we were going in the right direction. But whether the improvement was due to the increase in the number of CVV forms or to the diminution in the number of unnatural CCVs--or to something else--was not clear. (2) The CVC-type affixes were definitely most "valuable". They were better on the left of any joint than either CCVs (P about .01) or CVVs (P about .001), and probably also on the right in 3-termers, but this was less clear (P about .10). (3) CCVs were slightly preferred over CVVs in both medial and final positions but not significantly (P greater than .10). (4) Three-termers with two CVVs (medial and final) were definitely worse than ones with 1 or no CVVs (P about .001). (5) Half-reduced 2-termers with the long affix last (**menkatma**-type words) were greatly preferred over the other order (**kaplirdru**; P about .0001) and, indeed, over all other types of words taken together (P also about .0001).

On the basis of these in some ways puzzling results I decided to proceed along the AB strategy vector; as is explained in Sec. 6. But, clearly, a more precise componential analysis of the interactions of the affix forms with one another was going to be needed.

8. Taste Test #2: Intelligibility of Some Trial Words: This was an effort to investigate intelligibility by circulating a cassette on which 5 readers read a list of 90 "difficult words". The stimulus words had been selected from the trial words used in TT1 to be rich in intelligibility problems. They certainly were. However, structure was not varied systematically among them and it didn't work. There are only two results worth mentioning, one dolefully negative: (1) It was abundantly clear that we had a very large problem with consonant-clustering, especially at the C/CC-joint. The intelligibility of some consonants in some combinations at these joints was pretty bad. But that told us nothing about the combinations that weren't there. (2) Vowels, whether alone or in pairs, were distinctly more intelligible than consonants. I found this surprising and, later, useful. There were also some scattered results on the pronunciation of vowel-pairs; but these were not systematic enough to be useful. It was the failure of TT2 as an experiment but its success as a way-pointer which led to the more sharply focused work of TT3 on vowel-pairs and, later, of TT4 on consonant-joints.

9. Taste Test #3: Preferred Pronunciation of Vowel-Pairs: This was a very simple experiment that worked. TT3 was sent out with TL4/4 and presented 48 mock-Loglan words containing every possible vowel-pairing, except **ee**, in the CVV-terms of 2 kinds of 6-letter words: CVV+CCV and CVC+CVV. We received 51 responses; and the results, while hardly a consensus, were consistent phonologically. That is to say, a nearly-systematic phonological pattern emerged: (1) **ai ei ui** should be monosyllabic in both contexts (initial and final; both P's around .05). (2) **ae ea eo eu oa oe ou** should be disyllabic in both contexts (P's around .05). (3) **au ie iu oi ue** should be monosyllabic when final (P around .05) and probably also when initial but differences not significant (P greater than .05). (4) **aa ao oo** should be disyllabic when initial (P around .05) and probably also when final but not significant. (5) With **uo ua ia** the Ss favored monosyllables in both contexts but non-significantly. (6) With **ii** Ss favored the disyllable in both but non-significantly. (7) With **io** Ss favored the monosyllable when final, the disyllable when initial; but neither significantly. (8) On **uu** the Ss were split in both contexts.

The pronunciation system that completes this nearly-perfect pattern is, at the same time, one that meets the requirements of the new morphology: (a) all **i**-containing and **u**-initial pairs except **ii** and **uu** are to be regarded as monosyllables in both positions, or at least neither vowel in such a pair is to be given the word-stress when the pair is final

even when pronounced disyllabically (it is almost a trivial matter whether such pairs are pronounced disyllabically or as true monosyllables when they inhabit the penultimate term of a complex, provided that, if disyllabically, the second member of the pair is stressed); (b) **ao** is to be a monosyllable in both positions (because it is a true diphthong and can be, and is thus definitely distinguished from disyllabic **au**); and (c) all other vowel-pairs are to be treated as disyllables in both positions.

Word-stress can then be on the penultimate syllable, as usual. This is the system I have followed in contriving the pronunciation guides for the remade complexes, and in doing the resolutional analyses. And, with the single qualification that when a **u-** or **i-** initial pair follows either of the vocalics **r** or **l** in final position, as in **ka'krui,** when there is a strong temptation to treat this normally monosyllabic pair as a disyllable and let the stress slip rightward, producing **kakru'i,** the system seems to work. But even this is not a serious problem. If we felt that these "monosyllables" were contextually biphasic, and that in these post-**r/l** contexts they are "really disyllables", then all we have to do is exclude them from the ranks of of the vv's that make proper **spai**-form words when in such contexts. For if ***kru'i** is to be, like ***spe'a,** not acceptable as a CCvv-form borrowing, then there is no problem with the **-kru'i** part of **kakru'i** being heard as one. But this is an open issue. I seem to have no trouble saying **ka'krui;** and prefer to. And doing so retains **krui** among the **spai**-form borrowings. But perhaps most speakers will have trouble, or will not prefer to. In which case **rui** and kin will have to be reclassified as disyllables and some words remade. The problem evidently needs some further study.

10. Taste Test #4: The Intelligibility of the Consonant Joints: The stimulus materials for this important study, namely 739 mock-Loglan words of CV'CCV- and CV'CCCV-form built to exhibit every possible C/C and C/CC joint in the language, were prepared in the Spring of 1981 and distributed to a selected set of reader/listener-pairs. The purposes of the study were (i) to test Anthony Lovatt's 1977 hypothesis that many medial CC combinations proscribed in 1975 were actually quite intelligible, and (ii) to identify the most unintelligible combinations at the C/CC joint so as to avoid or hyphenate them in the construction of complex predicates. By randomly presenting both CV'CCV and CV'CCCV words on the same test, subjects were given an opportunity to mishear consonant doublets as triplets, and vice versa. Thus the accuracy of consonant-counting is also at issue here, a matter that will be crucial for distinguishing some primitives from some complexes in the new morphology.

Three pairs of loglanists, the Lees, the Prentices and the Jenners, responded; and my daughter Jenny and I made a fourth. Collectively, these four reader/listener pairs provided me with 7 massive "slabs" of data across the entire stimulus field, 739 responses in each slab. (Three of these couples exchanged roles and did the whole experiment twice...an experience that consumed part of many weekends. We owe a lot to these doughty subjects.) Understandably, the data took some time in coming in; but in late August and early September 1981 I was able to make the first of several analyses. The CC results seemed very firm; but I was forced into a conservative statistical posture on the C/CC joints by the small size of the sample: that is to say, of calling some C/CC joints "bad" that might not be bad given information from a larger sample. But the results were good enough to arm the word-making algorithm Bob McIvor was then preparing for TT5. This algorithm was instructed to insert hyphen **r** at the C/CC joints of all the D-Set trial words that Early TT4 would tentatively labeled "bad". The same words were also to be submitted to the TT5 subjects without hyphens; so a rough check on these early TT4 results was to be built into TT5.

But this was not enough. Intelligibility is essentially an acoustical phenomenon between a source and a receiver. It cannot be predicted, I was learning, by the subjective judgement of single individuals. (There is remarkably little correlation between joints that individuals think are bad and those that are bad dyadically.) So in February 1982 I sent out, with TL5/3, a second wave of TT4 forms. The response was gratifying. Some 37 loglanists managed to find a listener, or to play both roles by recording and listening to themselves, and so to respond to at least one of the 9 "partitions" (8 of 82 words and 1 of 83) of this immense block of test materials. As a result I obtained 4 1/3 "composite slabs" to be added to the 7 slabs produced by the first wave. (There were "4 1/3" additional slabs, and not some integral number, because, from the 2nd wave, I received 5 sets of responses to 3 of the 9 partitions but only 4 sets to

each of the other 6.) Thus, in the end I had 8376 intelligibility responses to analyze, 5173 of them provided by the 8 subjects of the 1st wave and the rest from the 37 subjects of the 2nd.

The findings from the 2nd wave confirmed nearly all the solid results from the 1st wave and allowed me to abandon the conservative posture I had taken toward the C/CC joints. It was now clear that some joints I had tentatively labeled "bad" were only difficult...not really unintelligible to my subjects. These findings are shown in Tables 1 and 2, in which the entries for the combinations I judge now to be truly unintelligible are in bold-face, and those for the merely difficult ones are underlined. You may wish to take a preliminary look at those tables. Both the tabled variables f and i are measures of unintelligibility, but slightly different ones as will now be explained.

The data from the two waves have been consolidated in the tables. It would have doubled their length to show the two waves separately, and there are no statistically significant differences between waves. Even the personal error-rates, which seemed certain to be different in the two waves--after all, some of the most experienced speakers and listeners in the whole of Loglandia were in the 1st one--turned out as a whole not to differ significantly between waves. People in the 1st wave had a 15.2% average error-rate; in the 2nd, 18.6%. This is a difference, and in the expected direction; but it is the sort of difference that could occur by chance alone about half the time (P = .48).

But error rates do differ significantly between individuals in the sample. Three individuals in the 1st wave tied for the lowest overall personal error-rate, which was 7% over the 9 partitions, and the best single performance on any one partition was 2.4% (2 errors in 82 tries; these are difficult words). In contrast, three individuals in the 2nd wave had personal error-rates in excess of 44%, and the highest rate observed for any one subject was 55% (45 errors in 82 tries). Yet, despite this concentration of the very low and the very high error-rates in the 1st and 2nd waves, respectively, there was, as I say, no difference between the two waves that could be believed in statistically. Obviously there were both some practiced speakers and some novices in both waves.

That there were nearly order-of-magnitude differences in personal error-rates among our subjects (7% vs. 55%) is an important fact in itself. Evidently even these often very difficult mock-Loglan words turned out to display vaguely familiar patterns to our most experienced loglanists. So the reasonable words on the list, though meaningless, were really quite easy for them to hear. On the other hand, the subjects I am calling "novices" (and one or two genuinely were) made all sorts of errors on words that are really quite plain to the loglanist's ear...if properly pronounced. So obviously either the readers or the listeners on these teams, and probably both, were thrown off their stride by the outlandishness (Loglandishness?) of the material, as the practiced listeners and readers most assuredly were not. So this grand, order-of-magnitude effect of what can only be inferred to be learning, on both the Loglan listening and speaking arts, ought to give some comfort to our novices. With sufficient practice, even totally unfamiliar Loglan words can evidently be plainly heard. No one doubted that, I know. But it is comforting to have some hard data on this point.

Just because there were such large individual differences between the novices and the practiced speakers in our sample, I was obliged statistically to take error-rate differences into account in analyzing the data. After all, if a subject made only one error in a column of 41 words, as often happened when one of our three best listeners was doing the listening (and his or her equally practiced partner, the reading), that error almost certainly meant much more about the unintelligibility of the word which provoked it than if that same error had been one of 20 made in the same column. (Exactly this kind of difference occurred among our subjects.) So in addition to the frequency with which errors were made on individual words, which is the upper figure in each cell in the tables, I also tabled the "error information" collected for that word from the error-rates of the individual subjects who missed it.

I defined the "error information" on a given word as the sum of the negative logarithms (Claude Shannon's measure) of the "local probabilities of error" being exhibited by the subjects who missed that word at the time they missed it, that is to say, of the probability of error in the column in which they missed it, multiplied by 10 for convenience in tabling. Thus, each error is regarded as a "message"; and the information content of that message is the familiar negative logarithm of the probability of its having been "sent". I called this sum of information i, and it is the lower figure in each cell of both tables.

TABLE 1. THE PHONOTACTICS OF THE C/C JOINT

16 X 17 matrix; 232 cells occupied. Upper figure is error frequency f,
1st C, row; 2nd C, column. lower, error information i; see text.

	/B	/D	/G	/V	/M	/N	/L	/R	/J	/Z	/P	/T	/K	/F	/C	/S	/H	Sum	Avg
B/	\\	-	2	2	1	-	1	-	**6.4**	1	1	2	-	-	2	-	3	21.4	1.3
	\\\	-	12	18	7	-	9	-	**54**	5	9	13	-	-	10	-	15	152	9.5
D/	-	\\\	-	1	1	2	1	-	2	-	-	-	-	1	1	3	1	13	.8
	-	\\\	-	6	4	10	8	-	8	-	-	-	-	6	5	27	6	80	5.0
G/	-	1	\\\	1	1	1	-	1	3.6	1	-	1	.8	1	-	2	1	15.4	1.0
	-	8	\\\	4	8	4	-	11	33	7	-	4	5	5	-	20	4	113	7.1
V/	.8	-	.8	\\\	1	-	1	-	2	3	1	.8	-	1	2	1	1	15.4	1.0
	3	-	8	\\	8	-	7	-	12	23	9	6	-	9	14	9	4	112	7.0
M/	1	-	2	2	\\	.8	1	1	-	2.8	3	5	2	.8	-	1	1	23.4	1.5
	4	-	25	14	\\\	3	5	8	-	19	24	37	13	3	-	4	6	165	10.3
N/	1	-	-	-	-	\\\	-	1	3.6	1	2.6	-	-	-	-	-	-	9.2	.6
	8	-	-	-	-	\\\	-	9	30	8	24	-	-	-	-	-	-	79	4.9
L/	-	-	-	-	-	-	\\\	-	-	2	-	-	-	1	-	-	-	3	.2
	-	-	-	-	-	-	\\\	-	-	13	-	-	-	4	-	-	-	17	1.1
R/	-	1	-	-	-	-	-	\\\	-	2	-	-	-	-	1	2.8	-	6.8	.4
	-	6	-	-	-	-	-	\\	-	9	-	-	-	-	5	16	-	36	2.3
J/	2.8	1	.8	2	1	2	1.8	-	\\\\\\\		-	4	-	3.6	\\\\\\\		3.6	22.6	1.7
	31	9	3	18	11	15	12	-	\\\\\\\		-	21	-	22	\\\\\\\		15	-157	12.1
Z/	.8	1	-	2.8	2.8	1	3.6	1	\\\\\\\		1	-	4	3	\\\\\\\		2	23.0	1.8
	6	5	-	25	21	4	15	4	\\\\\\\		13	-	25	23	\\\\\\\		16	157	12.1
P/	\\	-	-	3	-	1	-	2	\\\\\\\		\\	-	-	5	1	-	.8	12.8	1.0
	\\\	-	-	24	-	8	-	10	\\\\\\\		\\\	-	-	31	5	-	7	85	6.5
T/	-	\\\	2	2	1	1.8	1	1	\\\\\\\		1	\\\	1	1	-	2.8	-	14.6	1.1
	-	\\\	11	14	6	10	8	5	\\\\\\\		4	\\\	7	7	-	19	-	91	7.0
K/	-	-	\\\	1	-	2	-	-	\\\\\\\		1	2	\\\	3	-	2	2	13	1.0
	-	-	\\\	11	-	12	-	-	\\\\\\\		7	10	\\\	12	-	11	10	73	5.6
F/		-	1	\\\	1	1	1	-	\\\\\\\		-	1	2	\\\	2	2	2	13	1.1
		-	8	\\	7	5	4	-	\\\\\\\		-	7	11	\\	8	11	17	78	6.5
C/	.8	5	3.4	4	.8	-	-	1	\\\\\\\		3	1.6	-	2.6	\\\\\\\		3	25.2	1.9
	3	36	20	21	7	-	-	16	\\\\\\\		22	10	-	18	\\\\\\\		21	174	13.4
S/	**6**	5	2.6	3	-	-	-	-	\\\\\\\		-	1	1	1	\\\\\\\		1	20.6	1.6
	45	25	17	25	-	-	-	-	\\\\\\\		-	4	5	6	\\\\\\\		9	136	10.5
f:	13	14	15	24	11	13	11	8	18	13	14	18	11	24	9	17	21	252.4	
	1.0	1.0	1.0	1.7	.7	.8	.8	.5	2.2	1.6	.9	1.2	.7	1.6	.8	1.4	1.3	f' = 1.1	
i:	100	89	104	180	79	71	68	63	137	84	112	112	66	146	47	117	130	1705	
	7.7	6.4	7.4	13	5.3	4.7	4.5	4.2	17	11	7.5	7.5	4.4	9.7	3.9	9.8	8.1	i' = 7.4	

For example, in Table 1 on the C/C joints, B/J is apparently the "most unintelligible" combination. Evidently "6.4" subjects missed it; and that is the largest value of f that there is in this table. (Actually, 7 out of 12 subjects missed it. But because the sample size of 12 applies to only 3 out of the 9 partitions, and as exactly 11 subjects had responded to all the other words on the list, all data has been transformed to a (mythical) uniform sample size of 11. Thus 7/12 = 6.4/11, approximately.) And evidently those 6.4 subjects collectively contributed "54 units of information". We can easily convert that i-value back into the average of the local probabilities of error...well; into something pretty close to that average, anyway. Thus 54/6.4 = 8.4375 gives the average amount of information i each subject who missed the B/J word had supplied; and this corresponds (after dividing it by 10, negating it, and taking its antilogarithm) to a certain probability: thus P = antilog -.84375 = .143. The antilogarithm of the average information systematically underestimates the average of the true local probabilities...by a lot if they were few and very different, by very little if they were numerous and/or similar. But it is the best that we can do without going back to the original data each time; and the average we are going to compare it with also contains this bias. Evidently the average local error-rate of those who misheard the B/J joint was not less than, and probably pretty close to, 14.3%.

Now this lower bound on the average local error-rate of the subjects who misheard a certain word is sometimes very useful. It tells us whether the practiced speakers were joining in the fun. Evidently they were on the B/J word...at least to some extent. For the average local error-rate of all our subjects, whenever they missed any of the CVCCV-words covered by this table, is 21.1%. (That's the total sum of i divided by the sum of f, in the bottom righthand corner of Table 1, the result divided by 10, negated, and its antilogarithm taken. 21.1% is higher than the average column-by-column error-rate, which happens to be 16.4% for all columns, only because, in calculating the average local error-rate, a column which has many errors is counted many more times than a column that has few.) Now 14.3% is substantially lower than 21.1%. And the only way it can have been any lower than the table average is if some better-than-average listeners were joining in the error-making fun. Thus we can argue that B/J is probably genuinely difficult to understand. For although 7 out of 12 (6.4 out of 11) is not really a very high frequency of error--compared to some of the error-frequencies in Table 2, for example--some better-than-average-listeners were almost certainly involved in the mishearing of B/J. It is not, in short, entirely a novice's error that we are looking at in the B/J cell.

Average local error-rate--let us call it 'r'--is not, of course, the only, or even the best, measure of intelligibility (if it were, that is the variable I would have tabled). r is only one of 3 measures I found useful, the others being i and f. But r is the one that tells us what mixture of novices and/or experts were missing a given word. That is never all we need to know. For example, look at M/DJ in Table 2. Only 3 Ss missed this word; but they generated a lot of information doing it: i = 28. Evidently there were some (normally) pretty good listeners among these 3. Sure enough, r for these 3 Ss was 11.6%, about half the average for the C/CC table, which is 20.7%. Still, only 3 out of 11 subjects missed this word. So a good many normally poor listeners must have been hearing it correctly. The very fact that many poor listeners heard the M/DJ word correctly surely means that we are not justified in calling it "unintelligible"! So let us put the bad performance of (1 or some of) our good listeners on the M/DJ word down to momentary inattention, or to random noise. (Perhaps a dog barked. Or there were undetected transcription errors.)

On the other hand, we might well think it reasonable never to label a word unintelligible if its r is much greater than the table average; for that would mean that the good listeners were **not** mishearing it. But, luckily, such cases do not exist if f is at all high. There is only one word among the 20 or so with highest i and f values that gives even an average r...and that word happens to be the only one that 10 of our 11 subjects missed! The word is **sejtci** in the J/TC cell. Its r is 20.9%, practically identical to the table average of 20.7% So, of course the 10 out of 11 Ss who missed **sejtci** were collectively showing only "average skill" when they missed it. They were nearly all the subjects that there were! So the fact that the r of a cell is substantially equal to, or even a little greater than, the table average is not much use to us when f is high. And the stronger criterion that r be much higher than the table average is not met even among "bare-majority" words, i.e., those with f = 6. Apparently, then, some good listeners were joining in **every** majority that missed a word.

So what we want to label unintelligible are apparently words that meet either of two criteria: (i) either they have very high f, in which case r will automatically approach the table average; or (ii), with moderate values of f, we must have evidence that at least some considerable fraction of the practiced speakers were among those who missed the word. This will be true only when i is sufficient to yield an r that is substantially less than the table average. It turns out that all words that have both f of at least 6 and i of at least 40 meet one or the other of these two requirements. And these are the ones whose entries are in bold-face in the tables. As for the underlined words, these are all those joints that might have been taken as difficult on one sort of evidence but not the other, or were "close misses" on both.

Considerations of this kind guided my studies of intelligibility in these tables and led to the selections shown. Using the same scale of unintelligibility for both tables, I came to the conclusion that, in Table 1, only 2 of the 232 C/C combinations investigated had proved difficult enough to be called "unintelligible": the B/J case we have already discussed and, on somewhat weaker grounds, S/B. There is only one complex with a B/J-joint in the present list of remade complexes, namely **pebrjio**; and I have, as you see, hyphenated it. There are no S/B's. So even the 2 C/C joints I am proposing be proscribed are not really very troublesome...since they hardly ever occur among the tuned words.

There are 3 more medial pairs which are underlined in this table, namely M/T, C/D and S/D. These may well be difficult enough for some speakers or listeners to want eased with hyphens even if we are not always obliged to. I do not myself find words with these joints very difficult--certainly not unintelligible--and I have not hyphenated their occurrences in the present list of complexes. On the other hand, if anyone decides to say **recrdo'u** instead of **recdo'u**, or **dasrdo'u** instead of **dasdo'u**, I (and others, I am betting) will certainly understand da. M/T occurs in the tuned CPXs 5 times; C/D 3 times; and S/D 8 times. As far as I know none of these 5 pairs occurs as a medial consonant-pair in any primitive. Thus, Tony Lovatt's hypothesis about "freeing-up the medial consonant-pairs" appears to be working splendidly.

Turning to Table 2 on the C/CC-joint, the outcome is if possible even more auspicious for the new morphology. It is true that, on the same scale of unintelligibility, I find 19 C/CC cases that are in principle difficult enough either to hyphenate or to avoid altogether. These are, in the order of their unintelligibility, G/ZB, J/TC, J/DJ, M/ZB, N/DZ, D/CT, K/DZ, G/TS, C/DZ, S/VL, P/DZ, V/TS, N/DJ, D/CM, T/VL, D/TS, J/TS, J/VR and C/VL. But Nature seems to have chosen sides here. Words involving these 19 unintelligible triplets almost never occur! In fact, in the present list of some 2000 remade complexes, I find only 3 words that involve these now-proscribed combinations; and, of course, all of them, like **pebrjio**, are now hyphenated. They are **rinrdzo** (N/DZ), **hadrcme** (D/CM) and **lagrzbutau** (G/ZB). There are, of course, other occasions for the phonotactic hyphen, namely all the extensive cross-hatched areas of both Tables 1 and 2. But finding a total of 4 more occasions, out of 2000, on which the new morphology dictates a hyphen is not, I think, cause for the smiting of foreheads or the wringing of hands.

Do take a moment to notice G/ZB, our worst case (f = 9; i = 84; r = 11.8). The test word was **magzbo;** and **magzbo** comes as close as any word on the list to being utterly unpronounceable...despite the fact that the speaker **thinks** he's said it with great precision every time. (As you will!) Confident as our readers were, however, that they had actually spoken this word as written, only 2 of our 11 listeners managed to hear the **z** in it. All the others--which includes **all** our best listeners, this time--heard the word as **magsbo**...as I'll wager you will, too. In other words, here is a word where the speaker's conviction that da is speaking correctly is almost absolute. But the listener seems to be unable to find enough acoustic cues to the speaker's intended performance to be able to repeat it correctly in de's own head. N/DJ provides another instance of this sort of "self-destructing" word. The test word in this case was **sondji.** Again, the speaker always **thinks** da's said the **d** when da reads **sondji.** In fact, da has no difficulty whatever with its pronunciation. But the **d** drops out, as it were, on its way to the listener. And so what the listener hears is **sonji.** Even those listeners who wrote down 'sondji'--and so got it "right"--almost certainly heard /SONji/. That is why, since the morphological difference between **sondji** and **sonji** is crucial in the new morphology--the difference between a primitive and a complex--we **must** introduce a hyphen into the former (should such a word ever occur), thus both speaking and writing **sonrdji** when we have the complex in mind, and banish its unhyphenated form from the language altogether. For if we kept

TABLE 2. THE PHONOTACTICS OF THE C/CC JOINT - Part 1, The Left Half
(Middle consonants B through K)
16 X 36 = 576-cell matrix.

	/B-		/C-							/D-			/F-		/G-		/J-	/K-	
	/BL	/BR	/CK	/CL	/CM	/CN	/CP	/CR	/CT	/DJ	/DR	/DZ	/FL	/FR	/GL	/GR	/JM	/KL	/KR
B/	\\\\\\\		2	1	1	-	4	2	-	2	-	4.4	.8	1	.8	-	3.4	1	2
	\\\\\\\		18	6	5	-	39	14	-	14	-	42	8	13	3	-	21	10	14
D/	-	-	2	.8	6	3.4	4.4	-	8	\\\\\\\\\\\			1	-	-	-	1.8	2	1
	-	-	10	3	50	26	27	-	60	\\\\\\\\\\\			8	-	-	-	11	16	10
G/	1.8	2	2.4	1	.8	3	2	-	4	4	-	6	-	-	\\\\\\\		3.4	1.6	-
	13	10	22	5	3	17	8	-	34	24	-	35	-	-	\\\\\\\		20	12	-
V/	2	3	1.8	5	2.6	1	3.6	1	5	2	2	4.6	2	2	-	2	4	.8	-
	13	18	11	30	16	4	16	11	30	11	9	23	11	12	-	14	34	3	-
M/	2	-	-	-	-	.8	3	-	1.6	3	-	2.8	-	1.8	3.4	1.6	.8	1	2
	10	-	-	-	-	3	21	-	14	28	-	16	-	14	26	12	27	11	15
N/	3	1	2	-	2.6	1	4	1	3	6	1	8	1	-	-	-	2	.8	1.8
	17	11	14	-	17	7	25	9	17	50	11	64	8	-	-	-	13	5	12
L/	-	2	2	3.4	1	1	2.6	2	1.8	3.8	-	3	-	-	1.8	-	-	-	-
	-	10	11	25	6	7	11	14	10	28	-	27	-	-	10	-	-	-	-
R/	-	-	3	-	-	1.8	2	2	5	4.2	1	-	-	1	-	-	6	-	.8
	-	-	17	-	-	12	18	17	39	20	3	-	-	8	-	-	36	-	5
J/	5.4	4	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							8	2.6	5	3	3	2	5	\\\\	3.8	.8
	28	22	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							72	19	34	17	23	13	24	\\\\	31	3
			\\\\\\\\\\\\\\\\\\\\\\\\\\\\														\\\\		
Z/	3	1	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							5	1	4	2.6	2	1	3	\\\\	1	1
	12	6	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							48	4	17	19	17	4	21	\\\\	5	8
P/	\\\\\\\		6.2	.8	-	.8	3	1	2	3	4	6.2	1	2	3	2	\\\\	5	2
	\\\\\\\		35	6	-	3	15	5	19	24	21	51	8	10	16	10	\\\\	39	7
																	\\\\		
T/	3	-	3.6	1	-	1.8	2	2	2	\\\\\\\\\\\			.8	-	1	-	\\\\	1	5.6
	24	-	13	8	-	6	12	11	16	\\\\\\\\\\\			3	-	9	-	\\\\	4	46
K/	1	1	6	1	1.6	-	4	1	4	4	1.8	7.4	1	.8	\\\\\\\		\\\\	\\\\\\\	
	8	11	39	7	8	-	34	5	24	34	6	63	6	6	\\\\\\\		\\\\	\\\\\\\	
F/	1	-	1.8	2	2.6	.8	2	3	2	4.8	-	4	\\\\\\\		1	-	\\\\	.8	3.6
	8	-	9	9	9	3	13	12	12	35	-	27	\\\\\\\		8	-	\\\\	3	18
C/	1.8	2	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							4	1	7	2	2.6	1	4	\\\\	3	1
	9	16	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							38	6	49	14	20	6	23	\\\\	17	9
			\\\\\\\\\\\\\\\\\\\\\\\\\\\\														\\\\		
S/	1	3	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							6	1	3.8	1	1	4	2.6	\\\\	1	1
	21	22	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							39	7	26	6	6	28	18	\\\\	5	4
f:	25	19	33	16	18	15	37	15	38	60	15	66	16	17	19	20	21	23	23
	1.8	1.4	2.7	1.3	1.5	1.3	3.1	1.3	3.2	4.3	1.1	4.7	1.1	1.1	1.4	1.4	2.7	1.5	1.5
i:	163	126	199	99	114	88	239	98	275	465	86	474	108	129	123	122	162	161	151
	12	9	17	8.3	9.5	7.3	20	8.2	23	33	6.1	34	7.2	8.6	8.8	8.7	20	11	10

TABLE 2 (CONTINUED). Part 2, The Right Half
(Middle consonants M through Z)

468 cells occupied

/M-	/P-		/S-							/T-			/V-		/Z-			
/MR	/PL	/PR	/SK	/SL	/SM	/SN	/SP	/SR	/ST	/TC	/TR	/TS	/VR	/VL	/ZB	/ZV	Sum	Avg	
2	1.6	2.8	-	1	1	1	2	3	2	5.2	-	5	.8	3	3	3	61.8	1.8	B/
13	11	16	-	8	6	6	14	20	11	26	-	28	3	18	12	15	414	12.2	
1	-	1	2	1.6	-	1	2	1	-	4	1	**6**	1.6	2	4	2.8	61.4	1.9	D/
5	-	4	20	11	-	5	16	6	-	24	4	**48**	8	13	35	20	440	13.3	
.8	3	2	-	3	3	2	-	1.6	-	4.6	1.8	**7**	1	2.8	**9**	5	78.6	2.3	G/
5	19	14	-	15	22	14	-	8	-	40	14	**50**	4	18	**84**	32	542	15.9	
.8	2	2	.8	1	3	2	1	5	-	4	1	**6**	\\\\\\\		4.6	3	80.6	2.4	V/
3	10	9	6	13	22	17	11	39	-	27	7	**53**	\\\\\\\		33	21	537	15.8	
\\\	-	1	-	-	1	-	1	2	-	2	1	4.4	-	1	**8**	1.8	47	1.3	M/
\\\	-	8	-	-	7	-	7	8	-	11	8	30	-	5	**68**	12	361	10.3	
4	.8	.8	-	1	4	3.6	1.8	-	1	4	-	3	2.8	1	1	3	70	1.9	N/
28	5	5	-	5	25	30	9	-	6	22	-	20	16	7	10	15	483	13.4	
2	.8	-	2	1	2	-	-	2	-	3	-	3	1	-	5	5	51.2	1.4	L/
14	3	-	19	7	13	-	-	15	-	21	-	16	8	-	35	29	339	9.4	
1	-	-	2.6	1.8	1	3.2	1	1.8	-	1	-	1	-	1	5	4.6	51.8	1.4	R/
4	-	-	21	22	5	23	5	7	-	6	-	16	-	6	42	32	364	10.1	
2.6	4.6	2	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							**10**	2.6	**6**	**6**	4	\\\\\\\		80.4	4.2	J/
17	33	16	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							**68**	16	**48**	**46**	22	\\\\\\\		552	29.1	
4.4	1	2	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							5	3	4	2	1	\\\\\\\		47	2.5	Z/
18	4	8	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							35	23	24	13	6	\\\\\\\		292	15.4	
1	\\\\\\\		1	1	2	1	1	.8	1	4	.8	5.4	2	4.4	\\\\\\\		67.4	2.3	P/
7	\\\\\\\		9	6	13	9	8	3	11	24	3	37	23	33	\\\\\\\		455	15.7	
.8	2.8	1	1	-	1	3	2	1	2.8	\\\\\\\\\\\			2	**6**	\\\\\\\		47.2	1.7	T/
3	16	4	5	-	6	16	12	4	24	\\\\\\\\\\\			18	**49**	\\\\\\\		309	11.4	
1.6	1.8	-	3	1	1	2	2	1	1	4.4	1	5	1.8	5.6	\\\\\\\		66.8	2.3	K/
8	6	-	28	7	5	13	14	2	9	29	6	25	10	36	\\\\\\\		449	15.5	
1	1	-	3	1	-	.8	3	1	4.8	2	1	1	\\\\\\\	\\\\\\\	\\\\\\\		49	1.7	F/
8	8	-	18	13	-	5	16	3	42	10	13	7	\\\\\\\	\\\\\\\	\\\\\\\		309	10.7	
-	2	2	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							5.4	-	5.4	5	6	\\\\\\\		55.2	2.9	C/
-	11	12	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							40	-	37	31	**43**	\\\\\\\		381	20.1	
.8	-	1	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							3	-	5	4	6	\\\\\\\		45.2	2.4	S/
3	-	11	\\\\\\\\\\\\\\\\\\\\\\\\\\\\							22	-	31	34	**55**	\\\\\\\		338	17.8	
24	21	18	15	13	19	20	17	20	13	62	13	67	30	44	40	28	960.6		
1.6	1.4	1.2	1.3	1.1	1.6	1.6	1.4	1.7	1.1	4.1	.9	4.5	2.1	3.1	5.0	3.5	f' = 2.1		
136	126	107	126	107	124	138	112	115	103	405	94	470	214	311	319	176	6565		
9.1	8.4	7.1	11	8.9	10	12	9.3	9.6	8.6	27	6.3	31	15	22	40	22	i' = 14.0		

it, it would, willy-nilly, always turn itself into **sonji** by the time it arrived at the listener's ear. (That is why, incidentally, **sanpa djano** is **saa'dja** in the new lexicon and not ***sandja**...a word that would self-destruct. **Sa'nrdja** is an equally intact alternative; but the **r** costs more than the CVV, as we'll see from TT5.)

In sum, TT4 was a happy study. It confirmed Lovatt's hypothesis beyond all reasonable expectations, and it sharpened up his notions about medial CCs in the very few places where he was wrong. Even more importantly, it dispelled the widespread notion, which I once shared, that the C/CC joint--so frequently called for by the new morphology--is inherently awkward and "bad". It is...in certain combinations. But they happen to be combinations that our language simply does not use, or can easily be tuned to avoid. Apparently, nearly all the C/CC joints that "arise naturally" in Loglan are intelligible (once its affixes are properly tuned). And those few that are not intelligible, are so very few that the hyphens we add to the language on their account will be virtually invisible features of the language.

11. Taste Test #5: Preferred Sequences in Complex Predicates and Tuning the D-Set: In the background material I sent out with TT5, I said, after explaining the research model, 'If the model works, TT5 may very well be the final data-gathering effort of the GMR team.' Well; it did, and it is. The model not only worked, but the performance of the subjects who participated in it exceeded my most sanguine expectations. The Lobeast has given every sign of having functioned exactly like that beauty-detecting algorithm I invited it to imitate.

By way of brief review for readers who weren't among its subjects, TT5 was sent out in February 1982 to about 190 loglanists. The stimulus material sent each loglanist was a random 1/185th of an immense corpus of some 13,000 trial words that Bob McIvor and I had prepared over the preceding 6 months. The trial words were generated from two distinct affix sets (Sets C and D, this time) by a computer programmed to express 1954 distinct "complex concepts" (2- and 3-term metaphors) in just about every possible way given those affixes. The test concepts were thus a large subset of the 2262 complex predicates in the 1975 dictionary; and each concept was expressed, on the average, 6.6 different ways. Each loglanist received an average of 10.5 concepts, and so about 70 words. But each was asked to rate only the two best words for each concept.

I received 76 returns covering 802 concepts from 48 loglanists, 12 of whom had elected (at our invitation) to do more than one form...a negligible, or even desirable, bias in the sample. Five respondents (again, at our invitation) elected to rate all the 70-odd words on their forms; and from their 350-odd responses I was able to get reliable estimates of the average most-probable ratings of the unrated words on the other 71 forms. (They were "reliable estimates" only ex post facto, of course; that is, known to be so only because they later behaved so well statistically. I certainly did not know beforehand that they would.) This gave me some 1840 directly-rated words, but also some 3390 indirectly-rated ones, or 5227 ratings in all. This is not only a very large body of data, but, by design, the words rated covered a truly extraordinary range of affix types and sequences, often with sufficient numbers of even rare types to make a detailed componential analysis possible.

In the end, the statistical analysis of these 5227 numerical ratings allowed the selection and fine-tuning of an affix set capable of remaking the 2000-odd complex predicates in the 1975 dictionary in "the most satisfactory possible way". It did this by enabling me to quantify satisfactoriness. That is, it made possible the assignment of a numerical value, or score, to any word that could be made with these affixes in such a way that each score constituted a prediction of the degree of preference that particular word would enjoy over other equally possible (and so, scorable) alternatives for that same concept. It was then a simple matter, for any state of the lexicon, to sum the scores of all its complexes, and, by replacing some of them with words of higher score, or by moving affix-assignments around among the primitives so as to cause a net increase in the scores of the words they made, or, finally, by remaking some primitives so to make better affix-assignments possible, to progressively increase the sums of scores for successive states of the lexicon until, demonstrably, no further increase was possible...unless, of course, the remaking of even more primitives was still contemplated. But there came a time when the cost of remaking even the most promising primitive left with uncovered complexes was palpably--though not, I fear, exactly quantitatively--greater than the small net increase

in word-scores its remaking promised; and the procedure was terminated.

The procedure described in the last paragraph may be called "tuning". It was the business of TT5 (1) to select one of the two test affixes sets, C or D, for tuning, and (2) to provide the quantitative instrument--the "tuning tables"--with which to tune the affixes of the set selected. That so much could be accomplished by a single experiment seems extraordinary...and we were, in fact, extraordinarily lucky in TT5. Every important question I asked of it, save one, was answered; and the one unanswered question was itself a "lucky negative". (I had asked it to tell me whether "natural" CVCs, e.g., **mat** from **matma**, differed in preference value from "unnatural" ones, e.g., **mam** from **matma** and **rel** from **trelu**; and it couldn't tell me. Or rather what it did tell me was that the difference was probably a very small one, and in any case so small as not to be able to punch through the variance in what was in fact a very considerable number of relevant cases. I regarded this answer as an extremely fortunate one. For as TT1 so strongly suggested, CVC-form affixes are indeed most contributory to high scores. To be able to use the **mam**- and **rel**-versions of them freely practically doubles their availability. So you can see why I regarded this one negative finding as luckily so.)

There will not be space here to describe all the statistically significant results of TT5. It was as full of them as a Christmas pudding; I have seldom seen a richer body of data. But I will recount all of those that led to the choice of Set D for tuning, as well as those that contributed to the construction of the Tuning Tables shown in Table 3. As to the numbers in those tables, please read the note in the bottom righthand corner first. All tuning scores are expressed in this arithmetic. A "tuning-point" is thus one-tenth of an interval on the original 10-interval (11-point) rating-scale. And please note that a 3-term CPX with all its terms reduced--that is, with 3 short affixes--contributes, on the average, 1 1/2 times as many such points to the sum of scores for the lexicon of complexes (the number I have labored to increase) as a similarly reduced 2-term complex. Of course the subjects "didn't see it that way". But we must. Therefore the scores in the table for the "SSS's" are in the 70's and 80's, while the maximum value of an "SS" (prettier words, surely) is only 56. Please don't let this bother you. We are not trying to **reproduce** the ratings made by our subjects, but to use them to predict preference between comparable words. Since a 3-term metaphor can't be expressed in a 2-term word, it can hardly be relevant that our subjects would like it better if it were. Just as it is only marginally relevant to the tuning problem that our subjects like **menkatma**-type words best of all.

With these cautions, we are now ready to consider the results.

11.1 Set C vs. Set D: The words made with D Affixes were better than the words made with C Affixes on every possible comparison. If we look at the most highly-rated D-word and the similarly "best" C-word for each concept, when the two were different (613 cases), the best D's averaged 64.8 and the best C's 61.8 (P about .004). One-third of a scale-interval (3 tuning-points) is evidently a big difference, statistically. Even if we add in the 189 cases in which the best D and the best C word were identical, the shared word in general rating higher, the difference is between 67.3 and 65.0, about 1/4 of a scale- interval (2.3 points) and still significant (P about .01). That's considering only the 802 concepts on which we got subjects' ratings. But when the tuning tables were constructed, I also scored (by computer, fortunately) all the C- and D-Set words that had been made (before tuning the D words, obviously) for the entire test set of CPXs. (There were 1941 of them by that time; 13 of them having somehow got "lost" in the recesses of my computer.) The difference now was astronomical. Taking the highest-scoring word made by the original algorithm for each set for each concept, the average score--not a rating now, but as obtained from the tuning tables--for the best D-words was 52.5 vs. 40.3 for the best C-words, or 12.2 points difference. That's more than a full scale-interval. (The P for this result is so small as to be incalculable from my tables: less than .000,000,...). This is because, of course, D makes many fully-reduced short words possible; C, many fewer. And a word with an unreduced term gets penalized, of course, by the tuning tables...despite the fact that (let me repeat my caution) **some** unreduced words (certainly not all) are rated higher than any others. So, as far as ripeness-for-tuning is concerned, the issue, clearly, is coverage by short-affixes. C has only 70.0% coverage; D had 93.6% coverage even before tuning began, and has 95% now.

So there was no question but that it was the D-Set that should be tuned. But then

I'm afraid there never was. The virtue of including C-words in this experiment was largely statistical. They offered large contrasts with the more uniform D-words on almost every structural dimension of the complex word. And so they mobilized larger chunks of variance in the ratings to be ultimately accounted for by those structural differences than could ever have been produced by D-words alone. So it was having C along, I'm convinced, that contributed to the remarkable sensitivity of nearly all the statistical tests performed. Here, then, are some of the other results shown significant by those tests:-

11.2 Natural vs. Unnatural CCVs (0's vs. 1's): The prime dimension of change in the series of affix sets which I had constructed was getting rid of more and more "unnatural" CCVs, e.g, **fra** from **farfu** (which I came to call a "good unnatural") and **dzo** from **madzo** (a "bad"). (Recall that we started out with the "maximize the CCVs" hypothesis, which turned out to be the worst strategy of affix-assignment ever tried.) A "natural", of course, was like **cli** from **clika.** Had I been right in doing this? Set C was orthogonal to this movement, or even retrograde. It used many more unnatural CCVs, although its biggest difference from the D set is that it used no CVVs at all. But how did the CCVs fare? Pitting the unnaturals used by both affix sets against the naturals, there was 1/3 of a scale-point difference in favor of the naturals (3.4 tuning-points; P about .0004). Now this is not only a hugely significant finding statistically, that is, indubitably real, but it is an "across the board" result. In 18 paired categories of words, which differed pairwise only in the naturalness of their CCVs--20's vs. 21's, 00's vs. 01's, 8y0's vs. 8y1's, and so on, to use the symbols of the Tuning Tables (see the Key of Table 3)--a difference in this same direction and usually of about this same magnitude always obtained in the average ratings of the two categories of words being compared. There is no question, therefore, that unnatural CCVs are bad...and, moreover, consistently and everywhere bad. But how bad? What is the tradeoff? Well; we'll see that those 3.4 points of badness are worth about 1/4 of a hyphen (which is pretty bad) but nearly twice as much as a switch from a monosyllabic to a disyllabic CVV (which is bad, but not that bad).

What about the "good" vs. "bad" unnaturals? And there was even a third category that I looked at, the "specials" like **tci** from **titci** and **tco** from **totco.** These, I suspected, were actually pretty good...perhaps as good as naturals. But unfortunately there wasn't enough data to handle these subdivisions of unnaturalness, not even enough to give firm negative answers as in the case of the "unnatural CVCs". So I took my first 3.4 tuning-points in hand and went on to the next question.

11.3 Monosyllabic vs. Disyllabic CVVs (4's vs. 3's): Here I had 14 pairable categories of words to look at. The differences, of course, nearly always favored the monosyllabic forms, as I had confidently expected from TT3. But they were not so big as the cost of an unnatural CCV (only 2.0 tuning-points, on the average), nor so significant (P about .02), nor even quite so consistent. In fact, in the context of leading CCV, the disyllables fared slightly better than the monos...by about 1.3 tuning points. (This is not a significant difference, however; and I ignored it. Still, it's worth keeping in mind that **clise'a** (a mock-word) might just be slightly better than **cli'sei;** and is probably not worse.) In all 6 of the other paired-comparisons, however, monos fared from 4.2 to 1.2 points better than the disyllables. As I say, there is an average 2.0-point preference for the mono over the disyllable; and it is significant. So we have a second (nearly) across-the-board tuning factor to work into the Tuning Tables.

11.4 The Cost of Phonotactic Hyphens: To answer this question I found I had 26 pairable categories, 13 small categories of words bearing phonotactic hyphens, 13 much larger ones with words of matched structure but hyphen-free. (A phonotactic hyphen is one that follows a CVC-form affix and is called for by the "difficulty"--in this case, the conservatively-calculated difficulty--of the ensuing consonant joint.) Fortunately for getting a solid answer to this question, there were far more phonotactically-hyphenated words among the trial D-words submitted to the raters than there are now among the CPXs listed in this Notebook. (And that, of course, was a consequence of the conservative posture I had taken toward the C/CC-joint after the first TT4 results! As I said, TT5 was a lucky experiment from first to last.) The differences between hyphen-free and hyphenated words were quite large (on average, about 12 points), indubitably significant (P less than

.000,000,...), and thoroughly consistent (in the same hyphen-disfavoring direction in all pairs of categories that had enough hyphenated cases to be meaningful). So apparently it costs about 12 points in word-prettiness to hyphenate a joint. That's about 3 1/2 times the cost of an unnatural CCV and about 6 times the cost of a disyllabic CVV.

I used this cost-figure to good advantage in tuning the affix-set...mostly by getting rid of joints that needed hyphenation. But we must remember that our subjects, all loglanists, were being introduced in TT5 to a totally unexpected feature of the new morphology, one meant to deal with a problem that had not even been described. So probably 12 points overestimates the real, or lasting, aversion of lo loglenta to the phonotactic hyphen. **Mekrkiu** is just not that bad. (Hyphen R had emerged from some formal studies I had made the previous Winter; my results had not been bruited about our shores.) Still, I cannot imagine that phonotactic hyphens will ever be regarded as altogether lovely features of the language, can you? They might, however, subside into aesthetic neutrality. However that may be, armed with this large, if perhaps ephemeral, cost-figure, I made correspondingly large efforts to reduce their number in the present set of complexes...and succeeded.

11.5 Short-Long vs. Long-Short 2-Termers (SL vs. LS): TT1 had found that 8-letter words with unreduced final terms (SLs), were the most delicious words on its list, e.g., **menka'tma** = 'tomcat'. Did that once-surprising result hold up? It did. Short-longs with CVC initials were still the most highly-rated category, earning an average rating of 57.9. (Short-shorts of CVC+CCV-form, with the CCV natural, came in a pretty close second at 56.8, again as before. This is not the comparison we're interested in here. Even so, it is the remarkable carry-over of even these detailed results from TT1 into TT5--with mostly different words and largely different subjects (at least 42% different)--that is one of the most impressive features of these taste-tests. It means that these elusive aesthetic matters at are somehow very real.) This time the short-longs were astronomically preferred over both kinds of long-shorts (P less than .000,00...), those with 5-letter terms plus Hyphen N from the C-Set, e.g., **kanpi'nflo** (or **ka'npinflo**), and those with 4-letter terms plus Hyphen R from the D's, e.g., **ku'nerdui.** The mean value of an LS from the D-Set was 43.9; so there were apparently 14 tuning-points to be gained in shifting from an LS-form to an SL-form...a move, of course, which is seldom possible. Still, in a cascade of tuning moves involving several primitives and many affixes, these 14 points are occasionally among those earned. The average difference in the values tabled in the SL and LS subtables of Table 3 reflects this only occasionally useful tuning factor.

As between the C and D words, the 2 competing varieties of the long-short complex had mean ratings of 47.7 and 43.9, a difference in favor of the C-words with their additive Hyphen N that is itself highly significant (P about .0001). However, the choice between hyphenation systems is not a tuning issue but a design issue. There are three engineering considerations which I'll discuss in Sec. 12 that argue for the 4-letter-plus-hyphen pattern of non-final term reduction and only one--a non-engineering one--that argues against it: it looks odd. But even that has a possible solution; see Sec. 14.

The strong preference our subjects show for short-longs over long-shorts is not often a tuning issue, as I mentioned above; and it is never one in making a word from a given metaphor. After all, the long-shorts in the dictionary are so because there are no short-shorts for them; and if no short-short is possible, how can there be a short-long? Still, this apparently strong aesthetic (or is it semantic?) preference is, or should be, a factor in metaphor selection. A metaphor that will yield a **menka'tma**-type word is apparently greatly to be preferred over one that will insist on a **ku'nerdui**-form.

There is another and more useful tuning-factor to be found among these unreduced words. That is the definite advantage that the CVC-affix has over the CCV-affix (both varieties lumped) as the initial term in a short-long. The difference is 3.2 tuning points (P about .02); and that coupled with the weaker advantage that CVCCV-prims have over CCVCV-prims as the final terms in these words (1.4 points, but not significant) probably makes the family of words to which **menka'tma** belongs literally "the most gorgeous complex predicates in the language": average rating = 59.2. By the way, CCV forms enjoy no comparable advantage over CVV-forms as last terms in long-shorts. That is, **ku'nercli**-types are **not** nicer than **ku'nerdui**-types. In fact, if anything, the monosyllabic Cvv's, at least, get slightly better ratings than the CCVs in this position...perhaps because they reduce the consonant-burden around this joint. So the TT1 pattern of the CVCs

being best of all and the CCVs often being no better than the CVVs (and sometimes worse) is also holding up.

11.6 Preference Patterns Among the Short-Shorts (SSs): Please look at the subtable labeled SS in Table 3 for this one. An analysis of variance showed that the interactions between the affix-types of the first and final terms of these 6-letter words are significant when only the three main affix-types (CCV, CVC and CVV) are considered (P about .02). But so are the row and column effects (first and final affix) of these same affix-types taken separately (although more weakly; P about .05). Adding in the two across-the-board effects we have already identified--the effects of naturalness on the CCVs and of syllabicity on the CVVs--produces the numbers you see in the SS table after rounding. In short, what you are looking at in that table are the combined effects of four factors working separately, each factor having been independently established as a statistically reliable predictor of word-goodness. In a sense, the entire table is significant.

The results? The 3 prettiest types of short-shorts are without question the 00's (e.g., **mre'cli**) at 56 points, the 20's (e.g., **ma'mcli**) at 55, and the 24's (e.g., **ka'mbei**) at 54. With a 2-point spread, there is little to choose between them. They are all word-forms that our loglanists, on the average, apparently find, on the average, especially pleasing. (The double qualification is necessary, of course, because it is the average loglanist we are talking about as well as the average word in each form-class so-scored. Still some stable properties of both sets of objects seem to be punching through.)

We now drop 2 points to what turns out to be the average of the distribution of tuned 6-letter words: a score of 52. There are 2 forms that earned this average: the 21's (e.g., **du'rfro**) and the 01's (e.g., **gre'dru**, 'grease-do'). Then dropping below the average, we have, first, **mre'fua**, **mamse'a** and **tue'dji**-type words at 50, the **dru'mro**-type at 49, **saa'dja** at 48, **mrefo'a** at 47, the **dru'sei** and **tei'dru**-types at 46, and the least attractive of the short-shorts, apparently, are the **snali'i** and **mou'dru**-types at 44: affix sequences 13 and 31. Inevitably, given our tuning factors, the worst sequences combine the worst affixes: the disyllabic CVVs (3's) with the unnatural CCVs (1's) in both orders.

Now if your own ear tells you that most of the scores in this long series, as you moved from 56 to 44, were in some clear sense "deserved", then these words are probably pretty good representatives of their form-classes. (And you are probably a pretty typical denizen of the Lo-beast who told us what the form deserved.) But remember that there is some variance left over to bounce around within the form-classes. Not all 00's are equally "lovely", of course. The phonemes themselves add their bits. Nor or all 13's or 31's equally "ugly". But the remarkable thing about this table is that every number in it is supported by 3 or 4 independently significant factors. We cannot be sure that the interpolations between them are correct, of course. But apart from some extremely small, untabled interactions, they are bound to be approximately correct for this sample of judges...whose tastes, we have learned, were remarkably similar to those of the more numerous sample who were given the far uglier words of TT1 to judge. Apparently something is pretty stable about word-handsomeness: both in time, across different affix sets, and with substantially different subjects.

11.7 Preference Patterns in Reduced 3-Termers (SSSs): Unlike the SSs, in which the interactions between terms were stronger than the main effects of the first and final affixes taken separately, among the SSSs there are no measurable interactions between terms. Nearly all the variance in the ratings of SSSs is accounted for by adding up the effects of affix-type on each term taken separately. The middle term makes the greatest contribution to the score (P about .01), with both CVC and CCV forms in this position adding about 6.5 points more than Cvv-forms do. (Only monosyllables were allowed here...a restriction I have since lifted.) The last term has the next greatest effect on the rating of an SSS (P about .02), with final CCVs (both kinds lumped) yielding about 3.3 points more than final Cvv's, which are, in turn, about 4.4 points better than the disyllabic CV'Vs in this position...an effect more than twice as large, in the final terms of these 3-term words, as the general effect of syllabicity on a CVV. So the greater value of a Cvv over a CV'V as the final segment of an SSS, ordinarily worth 2 tuning points and here 4.4, has been handled directly by the table in the case of the these reduced 3-term words. Finally, the first term has the smallest effect on the rating (P about .05) with initial CVCs adding in about 3.7 points more than CCVs do in initial position. In addition to these

TABLE 3. TUNING TABLES
Word-Scores Used in Tuning the Affix Set

2-Termers:

SS	-0	-1	-3	-4
0-	56	52	47	50
1-	52	49	44	46
2-	55	52	50	54
3-	48	44		
4-	50	46		

Subtract 12 for phonotactic r between S's; subtract 6 if between an S & L. See Key to symbols below.

SL	-7v8
0-	32
1-	28
2-	33

LS	-0	-1	-3	-4
5x,6x,7y,8y-	20	16	17	19

3-Termers:

SSS	-S				SSL
SS-	-0	-1	-3	-4	-7v8
00-	82	79	73	75	52
01-	79	75	70	72	49
02-	81	78	75	77	52
04-	74	71	68	70	47
10-	79	76	70	72	50
11-	75	72	66	68	47
12-	78	74	71	73	49
14-	71	67	64	66	45
20-	85	81	76	78	53
21-	81	78	72	74	51
22-	85	81	76	78	53
(23-	76	72	68	70)**	-
24-	78	74	70	72	49

SLS	-L-	
S-S	-5v6x-	-7v8y-
0-0	53	53
0-1	51	51
0-3	48	-*
0-4	49	-
1-0	51	51
1-1	48	48
1-3	45	-
1-4	47	-
2-0	55	55
2-1	52	52
2-3	49	-
2-4	50	-

LSS	L-	
-SS	5v6x	7v8y
-00	55	55
-01	52	52
-03	49	-
-04	50	-
-10	52	52
-11	50	50
-13	46	-
-14	48	-
-20	54	54
-21	52	52
-23	49	-
-24	51	-
-40	50	50
-41	47	47
-43	45	-
-44	46	-

KEY

S = short affix
L = long affix
S's:
0 = nat'l CCV
1 = unnat'l CCV
2 = CVC
3 = disyl'c CVV
4 = monosyl'c CVV
L's:
5 = CCVC
6 = CVCC
7 = CCVCV
8 = CVCCV
Hyphens:
x = /r n l/
y = /n m/

v = 'or'

*'-'= not made.
**Added later.

SLL,LSL,LLS	S is in:		
S	1st	2nd	3rd
0	25	26	26
1	24	25	25
2	26	26	-
3	-	-	24
4	-	24	24

Tuning scores correspond to 10 X the original tastiness ratings. Thus '55' = '5.5' on the original 11-pt. scale. The correspondence makes sense, however, only when applied to 2-termers. Thus the average value of an S in both 2- and 3-termers in these tables is 24.82. So the tuning value of an average SSS is worth 150% of an average SS. The raters did not, of course, see things this way.

three main effects, the general 3.4-point superiority of the natural CCV over the unnatural one was also interpolated into the SSS subtable, after taking account its very different distribution among the cells of the SSS table.

The result? The 220's and the 200's are, at scores of 85, distinctly the most pleasant reduced 3-termers: say, **matsu'ncli** and **matsni'cli** (both concocted). At the other end, the 143's are predicted to be most disagreeable: say **snakiase'a** (again, concocted; but from real affixes). I suspect that this strong aversion to vowel-rich complexes among our loglanists may be temporary...more temporary, say, than the weaker but definite aversion they display to unnatural CCV's, which, involved in the decipherment task as they are, is likely to be a semantic matter, and so more stable.

In any case, I will let the reader work out--or find among the Remade Complexes--instances of the numerous intermediate scores in this large table. Other subtables in Table 3 are derivations of the main ones.

11.8 Tuning the D-Set and Measuring Coverage: The subtables of Table 3 provided the instrument by which the D-affixes were tuned into the shape in which you find them. There were, incidentally, four "tuning passes", with one exception of successively smaller net effects. On Pass #1, which was focussed on removing as many occasions for phonotactic hyphens as I could melt away--using the **new** TT4 tables, of course, to define the "bad joints" which call for them--I managed to add 1207 points to the 102,000 sum of scores with which I started: a 1.2% increase. On Pass #2, which was a struggle for coverage, I managed to add 452 more points: a 0.4% increase over the previous sum. On Pass # 3, only 146 points were added (I forget what I was doing, but it didn't work); only a 0.1% increase. And on Pass #4, on which I finally relented and remade some rather awkward primitives, I added 359 points, a 0.3% increase; and so ended with a total sum of 104,164 points: a 2.1% increase in the average tastiness of a word from the time tuning began. And I, at least, could taste the difference. The words really had become quite pretty to me during the course of all this massaging. But perhaps that is a phenomenon like the loved-one's face: you keep looking at it and it gets more beautiful, willy-nilly.

Coverage, which is quite a different matter, although related--and perhaps a more important measure of the success of the entire GMR project--increased from 93.6% to 95.0% during the course of the tuning operations. That amounts to a 1.5% improvement in coverage by the D-Set, and has brought the tuned D-Set to within a very small margin, at 95%, of matching the 97% coverage (by undecipherable affixes) found in the 1975 dictionary. (Coverage is measured over all the terms of all the complexes listed in that dictionary...except the **-sensi** words, which are not counted.)

Tuning reduced the initial coverage gap between the trial D-Set and the 1975 dictionary by about 40% (1.4/3.4). I do not think much more than that can be accomplished...except by remaking a very long series of additional primitives (which I do not want to do). For these would now be words of both low and diminishing power and so, would have small and diminishing effects. Still, the first 2 or 3 terms of this series might be worth tackling. (**Sange**, or one of its competitors, would be the first; **klini**, or one of its, the second. It is also possible that the **fasru/fasli** squabble over **fas** should be dissolved.) I am perfectly willing to remake any or all of these if the loglanists wish me to.

Probably no feature of any language has ever been engineered in quite this way, including computer ones. When I find myself doing this kind of thing I console myself with an observation of Francis Bacon's: "Things which have not been done cannot be done except by means that have never yet been tried."

And who knows? It just might fly.

12. 4-Letter vs. 5-Letter Non-Final Affixes: This is the issue, of course, that was presented in TT5 as a repeated choice between "Hyphen R" and "Hyphen N" words. But the morphological choice was really between the 4- and the 5-letter form of the non-final affix. There are three good engineering reasons why the 4-letter form will make a better language, and none that it will make a worse. Some you have no doubt already observed at work. But to review what I may have mentioned or you've observed:

(i) If built with 4-letter non-final affixes, the unreduced words of the language will, for the first time, be acoustically shorter than their metaphors. Thus, **sanprdja'no** is measurably more quickly spoken than **sa'npa dja'no**, with its 2 stressed syllables and an extra full vowel. To make Zipfean psychobiological sense to a speaker, the move from the

metaphor of a new concept to the first form of its complex **must** mean increased biological economy or da won't make it. Why should da? If we were to use, in contrast, the 5-letter form plus hyphen, getting **sanpardja'no** (**r** makes a better hyphen), the first complex would actually take longer than its metaphor to say. Not only that, but it looks longer and takes just that much more energy to write or type. So there is a definite loss of economy with the 5-letter form, and a distinct gain in economy with 4-letters.

(ii) The hyphen between the 4-letter term and the rest of the word looks and sounds like exactly what it is: a semantically negligible component of the word...a bit of morphological "glue". In contrast, the **par** (or the **pan**) in the hyphen-augmented right-half of the 6-letter form above looks like just another CVC-form affix. This is definitely misleading to the decipherer. It takes a double-take, and a good deal of intimate knowledge of the affixes, to see--or decide, finally--that it is not. This is a further loss of economy for the 5-letter plan: added decipherment time for that quasi-CVC.

(iii) The 4-letter form, while shorter than the original primitive, loses no information. After unpacking, no pair of primitives in the language (except the **spani/-spano** nationality/language pairs), will differ only in their last vowels. This is redundancy; and redundancy is fine so long as the primitives are moving about separately. In fact, we have needed more redundancy in the simple predicates of this language for a long time. After unpacking, we now have it. But when a primitive enters a CPX, becoming part of another word, redundancy is just what it can afford to lose. The 5-letter affix preserves the whole, often redundant original form. But it does so uselessly, and therefore wrongly, from an information-theoretic point of view. In contrast, the primitive being compressed into the 4-letter form seizes, in a manner of speaking, the opportunity given by its redundant vowel to drop some redundancy as it enters the more informative (because longer and less probable) context of the complex word. This is yet a third gain in biological economy for the 4-letter plan.

These arguments add up to something pretty decisive from an engineering point of view...no matter how our loglanists responded. Besides, there is good reason to expect that the aversive response they displayed toward 4-letter forms will turn out to have been ephemeral. Aversion is likely to diminish fairly rapidly as the engineering advantages of the 4-letter forms begin to impinge on their experience. After you have driven a high-performance car it doesn't matter that it is (was?) painted mauve.

Alright; what **was** wrong with "Hyphen R" and its associated 4-letter affix? What was mauve about it? Something pretty clearly was. The answer, almost certainly, is that it looked odd. In its neighborhood, it made the language look Polish. There are two possible solutions to this. One, we can wait for habituation to take over...wait for those consonant-surrounded letter 'r's to begin to look like the soft little sounds they are. Two, we can use another, or an alternative, letter. We'll consider our options in Sec. 14 on Consonant Buffering.

13. The Allomorphs of Hyphen R: Sometimes a hyphen is called for at a joint where one or both of the neighboring consonants is already **r**. In these contexts, **n** is used. So **n** is the "secondary allomorph" of the intraverbal hyphen. Such contexts are fairly rare. I haven't calculated the frequency of **n**-glued joints in the dictionary, but it just now took me five minutes to find one. Have a look at the Remade Complexes. You'll see that it takes some time to locate a hyphen **n** in a D-Set word. The one I found was in **letrnli'sta** = 'alphabet', an unreduced word. As you've heard (if you pronounced it), there's no difficulty with the pronunciation of this **-trnl-**. There is still that nice vocalic **r** sound--or that schwa--tucked away between the **t** and the **n**. Here's another: **spo'rncli**, 'spring-like', and **spo'rnvoi**, 'spring-jump', both rather pretty words. In these two words, the **r** is not vocalic; but even so, both the triplet and the quadruplet, led by **r** in each case, are remarkably easy (for us) to say.

What if the joint to be hyphenated has an **r** on one side and an **n** on the other? Then the tertiary allomorph **l** is used. I can't find one of these at all. And there may be, at the moment, none in the language. (One that did occur during tuning was so unlovely, however, that I remade a primitive to get rid of it.) These **-rln-** or **-nlr-** words may be difficult. Here's a concocted one: 'line-recline' (if that means anything) is **clina resto**, and that yields the unreduced form **clinlre'sto**. The **-nlr-** is definitely pronounceable; it is now the encased **l** that is vocalic. But some of the **-rln-** sequences, as I say, may verge on the monstrous.

Is any difficulty caused by the fact that **r** is also the primary allomorph of the hyphen used to attach certain privileged operators to CPXs? Not much. The hyphen here is part of some rather nice, preempted CVC-forms, all **r**-final, of course (see the CVC Assignment Table). For example, one is used in **nerdru**, 'one-do'. The fact that in **r**-followed contexts such **r**-final CVCs may be replaced with their **n**-final secondaries gives us a choice rather than a problem. The choice is the word-maker's, of course, or perhaps between CPX-making strategies in general. Take 'one-rule' as a relevant concoction. (**rui** is from **rulni**, once **gruni.**) **Ner** + **rui** won't work unless we doubly hyphenate it...as we may, of course, getting **ne'rnrui.** Or we may choose to make **nen** regularly available as an allomorph of **ner;** and then **ne'nrui** will work for 'one-rule'. Are problems caused by the fact that **nen** is also assigned to **nenri** (former **lenri**)? Not insurmountable ones. All it means is that, if we adopt such **n**-final secondaries, we give them privileges. Those privileges will be withdrawn from the primitives that just happen to have **n**-final affixes (the asterisked cases in the CVC Assignment Table). Thus, if **nen+r..** is always to mean 'one', then **nen** from **nenri** may never be used with following **r.** In particular, **nen** may not mean **nenri** in **ne'nrui.** Does this mean that **nenri rulni** ('in-rule') may never be expressed as a CPX? No; we can always hyphenate it as **ne'nlrui.**

So we can go two ways on this one. Either we can decide that all the **n**-final secondaries of the preempted CV**r**-forms (the asterisked cases) may never be used in the context of following **r;** or we can decide that the common ones like **nen** from **nenri** may be, and that in these cases--or in all such cases, for that matter--the operator-bearing CV**r**-form will carry a double hyphen. The consequences for this particular case would be that, on the first strategy, **ne rulni** would get **ne'nrui**, and that **nenri rulni** would get **ne'nlrui**...a bit awkward, perhaps. On the second strategy, 'one-rule' would be **ne'rnrui** and 'in-rule', **ne'nrui.** The choice seems clear here; but the matter can best be resolved by studying use-frequency over more cases...in fact, over all relevant cases. I invite some patient student of the new morphology to solve this problem for us. I would loan da my Eaton (a five-language word-frequency dictionary).

14. Consonant Buffering: We have all heard the Italian-American character in the movie say 'That'sa my boy.' Perhaps fewer of us have heard the Japanese person, faced with the same formidable consonant-clusters of English (**tsm** in this case), buffer them in approximately the same way...perhaps using a slightly different "buffering vowel". Loglan is a language meant to be usable, and so, speakable, by anyone on this planet. Inevitably, and wherever it indulges in them, its consonant-clusters will be buffered by those who cannot speak them any other way. I propose we welcome this event, and prepare for it.

There are now two stories about the sound-rhythms of Loglan...in fact, there always have been. In its strings of little words it is a smoothly alternating consonant/vowel, or consonant/vowel-group, language. But inside even its simple predicates there is always at least one consonant-pair, and in its complex ones, old or new, there are often vowel-singlets alternating with consonant-doublets, and even with occasional triplets. Probably the "consonant load" on the predicates of the new morphology is a little less than it was in the old one just because of the new CVV-form; but it could be made to be much less. (It would be much less, for example, if we used Cvv's wherever we could use them--see my comments among the Remade Complexes--which is something worth thinking about.) So it is in the predicates and the predicates alone that consonant-buffering will be needed, and will occur.

We now have a 6th vowel...actually, it's a 7th; we have always had vocalic **r** for use in names. Suppose we took the two vocalic allophones of our current letter 'r' and gave them to a 23rd letter. What "23rd letter"? What do we have left in the Latin (i.e., Western European) alphabet? After recently taking up 'h', we have 'q', 'w', 'x' and 'y' left. Two of these are strongly associated with consonantal sounds and could hardly be used to represent any vowel. (Imagine telling anyone that 'x' was a vowel!) But two of these letters, 'w' and 'y', often represent semi-vowels, and one of them, 'y', sometimes stands for a full vowel: for example, as in English 'happy'. That vowel (/i/, or /i/ in some dialects) is not either of our two homeless ones, to be sure, but it is a genuine vowel.

What would **mekykiu** look like to you? How would **kuncydui** and **rodjymadzysensy-madzo** look, to take some more formidable cases? If we understood--who are going to promulgate this orthography if anyone does--that either vocalic **r** or schwa may be heard wherever this new **y** phoneme occurs, that the British and Bostonians are going to be using

schwa for it, and the rest of us reading this Notebook are probably going to use **r**, then would it matter that we have "borrowed" a letter for this pair of sounds that never means either of them in any other language? Possibly not.

But there is a better plan. Why not adopt 'y' for schwa as an alternative to **r**, and continue to use 'r' for both its vocalic and consonantal values? I'm not sure that '**mekykiu**' looks any less odd than '**mekrkiu**'. One of them tempts me to use schwa, the other, vocalic **r**. If we had both letters in our alphabet, we could spell out this dialectical variation clearly. **Mekykiu** would be 'eye-doctor' in the British and Italian dialects of Loglan, and in many others; and **mekrkiu** would be a North American dialect-word for the same concept...just as **Pidr** and **Pitas** are different versions of 'Peter' now, and **Rl** and **Yl** may soon be American and British Earls.

So much is fun because dialects are fun. But what really matters is that we would then be prepared to take consonant-buffering nakedly and boldly into our language. Not in some second-rate, essentially unspellable version of Loglan, as Lower East Side Manhattan is "unspellable American", but in mutually intelligible and mutually legible dialects of a language that lived as happily on one side of an ocean as on another. For now either of these sounds could be used for consonant-buffering, either to buffer the joints that are "difficult" in every dialect--our "phonotactic hyphens"--or gratuitously and dialectically to ease the consonants at any joint-like place in any word, whether it was complex or not. Thus we could have **matyma** from Japan, perhaps, and **matrma** being used someplace else. And both words would be as good as **matma** is for all our mothers. And if a Japanese loglanist chose to write our word **sporncli** ('spring-like') as **sypornycli**, or even as **syporynycyli**, and speak the latter as /syPORynycyli/, wouldn't we understand da? In both speech and writing? (Sequences of schwa-syllables are certainly not uncommon in English speech.../FORtcynytli/, using this handy new schwa.)

Such clearly identifiable audiovisual buffers, however sprinkled between the consonants of the language, would give neither the resolver nor the lexer the least pause. For both **y** and interconsonantal **r** would then be "dummy phonemes": the ones that don't count morphologically. And the ones you forget about when you're looking things up.

It just might fly.

15. What You Can Do: You can let me know whether you think I ought to carry on re-making prims, about how many, and in fact which ones. Do any unserved ones strike you as especially deserving? You can also let me know if any of the new prims bother you. Do you think any should be re-remade? Or even un-remade? And please identify any CPXs in the present list that strike you as especially ugly. Suggest alternative ways of making them if you can. And if you see a way of improving scores by shifting affixes around, just let me know. (Of course you can also let me know the things you like about the affix set as well.)

Please put all your suggestions on separate 3x5 index cards. I'll execute any clear improvements and act on any consensuses that emerge. Until the new dictionary goes to press, we can make any changes we like in our word-making tools, primitives and affixes alike. But after that? They're likely to freeze. So whatever polishing we're going to do had better be done now.

You can also volunteer to be one of the word-makers on a "shakedown cruise" I plan to take with the new affixes during the next six months: grinding out the **next** 2000 concepts on the Eaton List. (It was tuned on one set of 2000; it should be tested on another...to get any oddities from the first set shaken out.) Doubling our CPXs will also enrich our dictionary, of course. But that's less important than sending the affixes themselves out into the world of meanings they'll eventually have to deal with. So if anyone wants a share of those 2000, let me know. The work will have to be coordinated, of course.

There are many other things, of course, that an individual can do. You can extend and prove the set of 1 mod 3 residuals, for instance. Or you can study the CV**n**-before-**r** problem exhaustively. But these are the main ones: to help me put a final polish on the affix-set, and to shake them down by sailing through the next 2000 concepts with them.

1797 Terms Covered

CVC-ASSIGNMENTS - Part 1: BAB to LUL

Shows Number of Terms Covered by Each Affix
(...) = later assignment. [...] = Preempted.

		--B	--C	--D	--F	--G	--J	--K	--L
B-	A	-	BAlCi 2	BADlo 1	-	-	BAdJo 1	BAKto 2	BALpi 2
	E	-	-	BEDpu 3	-	BEGco 2	-	BEKti 3	BEkLi 1
	I	-	BIlCa 7	BIvDu 1	-	-	-	-	BILti 4
	O	-	-	-	-	-	-	-	-
	U	BUlBi 1	-	-	BUFpo 1	-	-	BUKcu 4	BULju 2
C	A	CABro 2	-	-	-	-	-	tCAKu 2	tCALi 3
	E	-	-	CEDzu 2	CEFli 3	-	-	-	CELna 1
	I	-	-	CIDja 3	-	-	-	CIKtu 2	CItLu 1
	O	-	-	-	-	-	-	-	COLku 1
	U	-	CUtCi 5	-	-	-	-	-	-
D	A	-	-	-	-	-	-	DAKli 3	-
	E	-	-	DEDjo 4	-	-	DEdJo 1	[DEKto]	DEnLi 3
	I	-	(DItCa)	-	-	-	-	-	-
	O	-	-	-	-	-	-	-	-
	U	-	-	-	-	-	-	-	-
F	A	FAlBa 1	-	FAnDo 7	-	FAGro 5	-	-	FALji 6
	E	-	(FErCi)	FElDa 2	-	-	-	FEKto 2	FELda 1
	I	FIBru 2	-	FIzDi 2	-	-	-	FIKco 1	FILmo 10
	O	-	-	FOlDi 1	-	-	-	-	FOLma 3
	U	-	FUtCi 2	-	-	-	-	-	-
G	A	-	GAnCu 3	GAnDi 4	-	-	-	-	GAnLi 1
	E	-	-	-	-	-	-	-	-
	I	-	-	-	-	[GIGdo]	-	-	-
	O	-	(GOtCa)	-	-	-	-	GOKru 1	-
	U	GUdBi 2	-	GUDbi 16	-	-	-	-	-
H	A	-	-	HArDu 4	HAsFa 2	-	-	-	-
	E	HErBa 5	-	HEDto 4	-	-	-	[HEKto]	HELba 2
	I	-	-	HIDro 1	-	-	-	-	-
	O	-	-	-	-	-	-	-	HOLdu 4
	U	-	-	-	-	-	-	-	-
J	A	-	-	-	-	JAGlo 8	-	-	dJALe 1
	E	-	-	-	-	-	-	-	dJELa 2
	I	-	-	-	-	-	-	-	-
	O	-	-	-	-	-	-	JOKla 1	-
	U	-	-	-	-	JUGra 2	-	-	dJULa 3
K	A	-	KAnCe 1	KAmDa 1	KAsFa 5	-	-	KAKto 18	KApLi 17
	E	-	KECri 4	-	-	-	KErJu 1	-	-
	I	-	KICmu 1	-	-	-	-	KInKu 1	[KILto]
	O	-	KOrCe 4	-	KOmFu 5	-	KOrJi 2	KOKfa 2	KOLro 4
	U	KUBra 2	KUnC1 1	-	-	-	-	KUKra 9	-
L	A	bLABi 2	-	pLADo 1	-	LAnGa 9	-	LAKse 3	LALdo 6
	E	-	-	LEDri 1	-	-	-	LEnKi 2	LELpi 6
	I	-	LInCo 1	LIDji 8	cLIFe 1	-	LImJi 1	LIKta 1	LILfa 13
	O	-	-	LODji 5	-	-	-	-	-
	U	-	-	-	-	-	-	-	LUnLi 1

CVC-ASSIGNMENTS - Part 2: BAM to LUZ

1360 Possible
476 Assigned
35% Used

Coverage Figures are Before Tuning

*Must not be used with following r.

--M	--N	--P	--R	--S	--T	--V	--Z		
BAlMa 2	BANci 2	BAtPi 2	BARma 4	BASni 1	BATmi 1	-	-	A	
-	BENdu 2	-	BERti 3	-	BETcu 4	-	-	E	
-	-	-	---	BISli 1	BITsa 3	BIVdu 6	-	I	B-
BONgu 1	-	-	BORku 1	-	BOTsu 3	-	-	O	
BUNbo 1	-	-	-	BUSte 1	-	-	-	U	
-	CANse 8	CAPri 2	CARbo 3	CASlo 2	CArTu 1	CAVle 3	-	A	
-	CENja 6	-	(CERsi)	-	CETlo 4	-	-	E	
CIMra 3	CINta 5	-	CIRna 2	-	CIsTi 6	-	-	I	C-
COMtu 1	CONdi 6	-	CORta 9	-	COmTu 2	-	-	O	
-	-	CUPri 1	CURca 4	CUtSe 2	CUTri 9	-	-	U	
DAMni 6	DANci 8	DAsPa 1	DARli 3	DASpa 2	DAnTe 3	-	-	A	
-	DENro 2	-	DEtRa 3	[DESti]	-	-	-	E	
-	-	DIPri 2	DIRco 5	-	DITka 1	-	-	I	D-
-	DONsu 2	-	DORja 8	-	DOTra 3	-	-	O	
-	DUrNa 1	DUPma 2	DURzo 18	DUStu 2	-	DUVri 2	-	U	
FAMji 2	FANra 3	-	FARfu 7	FASru 3	FATru 4	FAnVe 17	-	A	
FEMdi 8	*FErNu 2	-	[FEr]	FESti 1	FErTi 2	-	-	E	
-	-	FItPi 3	FIRpa 3	-	FITpi 3	-	FIZdi 2	I	F-
FOrMa 22	-	-	[FOr]	FOSli 6	FOTli 10	-	-	O	
FUMna 2	-	-	[FUr]	-	FUTci 10	-	-	U	
-	GANta 26	-	GARni 23	GASno 1	GArTi 1	-	-	A	
-	GENza 6	-	-	-	GETsi 1	-	-	E	
-	GINru 1	-	-	-	-	-	-	I	G-
GOMni 1	-	-	GOtRi 2	GOtSo 3	GOTso 11	-	-	O	
-	GUNti 10	-	GUtRa 2	GUSto 2	GUsTo 2	-	-	U	
HArMo 3	HANco 9	HAPci 10	HARko 1	HASfa 7	HATro 4	-	-	A	
-	-	-	HERfa 1	-	-	-	-	E	
-	-	-	HIRti 2	-	-	-	-	I	H-
HOMpi 2	-	-	HORma 3	HOSpa 1	-	-	-	O	
HUMni 4	HUmNi 3	-	-	-	HUTri 1	-	-	U	
-	JANto 3	-	JAnRo 2	-	-	-	-	A	
-	-	-	-	-	dJETa 5	-	-	E	
-	-	dJIPo 3	-	-	dJITu 4	-	-	I	J-
-	-	-	JORta 1	dJOSo 2	-	-	-	O	
-	*JUNti 3	dJUPo 1	[JUr]	-	-	-	-	U	
KAMla 5	KANmo 1	KAPni 7	KAmRa 4	KArSa 2	KATli 8	-	-	A	
KEMdi 4	KENti 4	-	KERti 4	-	KETli 1	-	-	E	
-	KINci 12	-	-	-	-	-	sKIZo 2	I	K-
KOMcu 1	KONte 2	KOPca 3	KORti 9	KOnSu 1	-	KOrVa 4	-	O	
KUMtu 4	KUrNi 3	KUPta 3	KURfa 6	KUSmo 5	-	KUVga 4	-	U	
-	LANdi 5	-	LARte 5	LASti 1	(LATci)	-	-	A	
-	LENgu 3	LElPi 1	LERci 1	[LESta]	LETci 3	-	LEdZo 1	E	
LIMji 2	cLINa 8	-	fLORa 3	LISta 1	LITla 3	cLIVi 21	-	I	L-
-	-	-	-	[LUSta]	[LUsTa]	-	-	O	
pLUMa 2	(LUNla)	-	-	-	-	-	-	U	

CVC-ASSIGNMENTS - Part 3: MAB to ZUL

		--B	--C	--D	--F	--G	--J	--K	--L
M-	A	-	-	MADzo 14	-	-	-	MAnKo 4	MALbi 2
	E	-	-	-	-	[MEGdo]	-	MEnKi 4	MErLi 3
	I	-	-	MIDju 8	-	-	MIdJu 1	[MIKri]	[MILti]
	O	-	MOtCi 1	-	-	-	-	-	-
	U	MUBre 4	MUtCe 10	-	-	-	-	MUrKi 2	-
N-	A	NABle 3	-	NArDu 3	-	-	NAJda 1	-	NALdi 2
	E	NErBi 3	-	NEDza 5	-	NEGda 2	(NErJi)	-	-
	I	-	sNICe 2	(NIrDa)	-	NIGro 1	-	NIKri 1	NIrLi 4
	O	(NOtBi)	-	NOrDi 2	-	-	(NOrJi)	-	-
	U	-	-	-	-	-	-	-	-
P-	A	-	PAtCe 2	PAzDa 2	-	-	-	PAsKo 3	sPALi 6
	E	sPEBi 6	PEtCi 2	PEnDi 1	-	-	PEnJa 2	-	-
	I	-	PInCa 1	PInDa 1	PIFno 3	-	-	[PIKti]	PILno 2
	O	-	-	POlDi 1	-	-	POrJu 1	-	POLdi 10
	U	PUBli 1	PUCto 1	PUDru 2	PUrFe 3	-	-	-	PULso 1
R-	A	-	tRACi 3	RAnDe 6	-	fRAGu 3	kRAJu 2	dRAKa 2	pRALi 7
	E	bREBa 2	REtCa 6	REDro 1	REsFu 3	-	-	bREKo 1	tRELu 1
	I	-	RItCo 1	RIDle 1	-	bRIGa 2	-	bRIKi 1	bRILi 3
	O	-	bROCu 2	bRODa 3	ROFsu 4	-	ROdJa 7	-	ROLgu 7
	U	-	-	-	tRUFa 1	-	-	-	RULni 7
S-	A	-	SAtCi 11	-	SAlFa 3	-	-	SAKli 2	SALdi 8
	E	-	SEkCi 2	-	-	-	-	SEKta 2	SELji 20
	I	SImBa 1	SItCi 4	-	SItFa 3	-	-	-	SILtu 2
	O	-	-	SOlDa 2	-	SOrGu 2	-	-	SOcLi 11
	U	-	-	SUnDi 2	-	-	-	-	SULba 2
T-	A	-	-	-	-	TArGo 3	-	TAKna 13	TArLe 2
	E	-	TEtCu 2	TEDji 4	-	-	-	-	TErLa 3
	I	-	TIrCa 2	TIDjo 2	-	-	-	-	TrILi 1
	O	TOBme 3	-	-	-	TOGri 2	-	TOmKi 2	TrOLi 1
	U	TUBli 2	-	-	-	-	-	-	TUgLe 4
V-	A	-	-	VAlDa 1	-	-	-	-	VAtLi 8
	E	-	-	VEDma 4	-	-	-	-	-
	I	-	-	VIDre 3	-	-	VIdJu 3	VIzKa 2	-
	O	-	-	-	-	-	-	-	VOLsa 3
	U	-	-	-	-	-	-	-	-
Z-	A	-	-	-	-	-	-	-	-
	E	-	-	-	-	-	-	-	-
	I	-	-	-	-	-	-	-	-
	O	-	-	-	-	-	-	-	-
	U	-	-	-	-	-	-	-	-

Number of Affixes: Number of Terms Covered

--B	--C	--D	--F	--G	--J	--K	--L
17: 42	27: 83	41:139	13: 37	12: 41	12: 23	30: 96	50:218

CVC-ASSIGNMENTS - Part 4: MAM to ZUZ

--M	--N	--P	--R	--S	--T	--V	--Z		
MAtMa 2	MAlNa 1	-	MARka 2	-	MATma 5	-	-	A	
-	MENdi 8	-	MERji 10	-	METli 2	-	-	E	
-	sMINa 2	MIPli 1	[MIRdo]	MIkSa 3	MITro 1	-	-	I	M-
-	MONca 3	-	MORto 12	-	-	MOdVi 4	MOnZa 2	O	
-	-	-	MURsi 4	MUSlo 2	MUTce 2	MUVdo 9	MUZgi 5	U	
NAMci 7	[NANti]	-	NAtRa 2	-	NATli 3	-	NAdZo 7	A	
-	*NENri 17	-	[NEr]	NESta 1	NETre 2	NErVi 2	-	E	
NIMla 8	*NIrNe 2	-	[NIr]	-	NITci 1	-	-	I	N-
NOrMa 1	-	-	[NOr]	-	NOTbi 13	-	-	O	
NUMcu 8	-	-	[NUr]	-	-	-	NUZvo 3	U	
-	PANba 2	-	PARti 10	PASko 10	PATpe 1	-	PAdZi 1	A	
-	PENso 6	-	PERnu 11	-	PEnTa 7	-	-	E	
-	PINti 4	-	[PIr]	PISmi 3	-	-	-	I	P-
-	PONsu 14	-	PORli 6	-	POsTa 4	-	POZfa 2	O	
PUbMu 1	PUNtu 5	-	PURda 10	-	sPUTa 3	-	-	U	
fRAMa 1	*RANta 2	-	[RAr]	RASto 1	pRATi 2	-	RAZnu 3	A	
fREMi 1	*RENro 4	REtPi 3	[REr]	RESfu 6	REsTo 1	REVri 1	fREZi 4	E	
tRIMe 2	RINta 2	-	RIRda 2	RISpe 1	-	-	pRIZi 1	I	R-
-	-	-	[ROr]	ROfSu 1	-	-	mROZa 1	O	
-	gRUNu 2	-	-	pRUSa 1	RUTma 2	-	-	U	
SAMto 3	SANpa 9	SAPla 1	SARni 3	SAnSe 6	SATro 1	-	-	A	
-	*[SENti]	-	[SEr]	SEnSi 24	SETci 2	-	-	E	
SIMci 1	SINma 4	-	SIRna 3	SISto 5	SITfa 6	-	SIdZa 1	I	S-
SOrMe 2	*SONda 7	-	[SOr]	-	SOlTe 1	-	-	O	
SUMji 3	SUNho 3	SUPta 1	SURla 2	-	SUTme 3	SUrVa 5	-	U	
TArMu 7	TrANa 2	-	TARci 4	-	-	-	-	A	
-	*TENri 1	TEPli 5	[TEr]	-	TETri 3	-	-	E	
-	TINmo 1	-	sTIRe 2	-	-	-	-	I	T-
-	*TOrNi 1	-	[TOr]	-	TOTnu 2	TOVru 3	-	O	
-	-	-	TURka 9	-	-	-	-	U	
-	VAlNa 4	VAlPu 2	(VApRo)	-	-	-	-	A	
-	*VENdu 4	-	[VEr]	-	VETci 4	-	-	E	
-	VINjo 7	-	VIRta 3	VIrSa 1	-	-	VIZka 8	I	V-
-	-	-	[VOr]	-	VOlTi 1	-	-	O	
-	-	-	-	-	-	-	-	U	
-	ZAvNo 2	-	-	dZASo 2	-	ZAVlo 16	-	A	
-	-	-	-	-	-	-	-	E	
-	-	-	-	-	-	-	-	I	Z-
-	-	-	-	-	-	-	-	O	
-	-	-	-	-	-	-	-	U	

--M	--N	--P	--R	--S	--T	--V	--Z	Totals:
34:123	55:265	20: 54	47:237	36:114	53:185	14: 97	15: 43	476: 1797

AFFIX ASSIGNMENT

By Monos: 950
By Di's: 470
Terms: 1420

CVV-ASSIGNMENTS - Part 1: BAA to LUU

1420 Terms Covered; 67% by Monosyllables

		--A		--E		--I		--O		--U	
B-	A	BArmA	1	(BAnsE)		BAlcI	3	BAksO	6	-	
	E	-		-		BErtI	10	BEgcO	10	BEndU	2
	I	BItsA	4	BIdjE	5	BIltI	4	-		BIvdU	6
	O	-		-		BOtcI	3	-		BOtsU	5
	U	-		BUstE	3	BUlbI	1	-		BUkcU	2
C-	A	-		CAnsE	2	CAnlI	21	CAbrO	1	-	
	E	CEnjA	57	-		CErsI	3	(CEtlO)		(CEdzU)	
	I	CIrnA	2	-		-		CIrzO	1	CIktU	3
	O	COrtA	1	-		-		-		COmtU	1
	U	(CUrcA)		CUtsE	42	CUtrI	3	CUndO	1	-	
D-	A	DAnzA	5	-		DAncI	2	DArtO	1	-	
	E	DEtrA	7	-		DEnlI	9	(DEdjO)		DErtU	1
	I	DItcA	5	-		DIsrI	3	DIrcO	12	DIslU	4
	O	(DOrjA)		-		-		-		DOnsU	32
	U	DUrnA	6	-		DUvrI	7	-		-	
F-	A	FAlbA	1	(FAnvE)		FAljI	1	FAndO	3	FAtrU	2
	E	FEldA	1	-		FErcI	1	FEktO	2	FErnU	1
	I	FIrpA	4	-		FIldI	1	FIlmO	4	-	
	O	FOlmA	66	-		FOtlI	1	-		-	
	U	FUmnA	10	-		FUtcI	6	-		-	
G-	A	GAntA	1	-		GArnI	4	-		GAncU	4
	E	GEnzA	1	-		GEtsI	2	-		-	
	I	-		-		-		-		-	
	O	GOtcA	2	-		-		GOtsO	40	-	
	U	-		-		GUntI	5	-		-	
H-	A	HAsfA	11	-		HApcI	2	HArkO	3	(HArdU)	
	E	(HElbA)		-		-		-		-	
	I	HIjrA	1	-		-		-		-	
	O	HOrmA	3	-		(HOmpI)		-		HOldU	3
	U	-		-		HUtrI	3	-		-	
J-	A	-		-		JAltI	3	JAglO	5	-	
	E	-		-		-		-		-	
	I	-		-		-		*dJIpO	2	*dJItU	1
	O	JOrtA	2	-		-		-		-	
	U	JUgrA	3	-		JUpnI	2	*dJUpO	4	-	
K-	A	KAmlA	10	KAncE	1	KAtlI	17	KAktO	25	KAngU	2
	E	*cKElA	2	-		KEcrI	2	(*cKEmO)		KErjU	4
	I	-		-		KIncI	1	-		KIcmU	1
	O	KOrvA	1	KOrcE	1	KOrjI	7	KOlrO	5	KOmfU	1
	U	KUvgA	14	-		KUncI	16	KUsmO	1	KUmtU	2
L-	A	LAngA	1	LArtE	5	LAndI	7	-		-	
	E	LEtrA	1	-		LEtcI	16	LEnzO	1	-	
	I	*cLInA	4	-		*cLIvI	6	LIkrO	1	LItnU	5
	O	*fLOrA	1	-		LOktI	1	-		-	
	U	*pLUmA	1	-		-		-		-	

AFFIX ASSIGNMENT

CVV-ASSIGNMENTS - Part 2: MAA to ZUU

425 Possible
216 Assigned
51% Used

*Medial derivation (cCVcV).

		--A		--E		--I		--O		--U	
M-	A	MAtmA	2	-		MAtcI	9	MAdzO	263	-	
	E	MEnsA	1	-		MErlI	9	(MEtrO)		-	
	I	MIlfA	5	-		MIplI	1	MItrO	3	MInkU	2
	O	MOncA	3	-		MOdvI	3	-		-	
	U	-		MUtcE	1	MUzgI	2	MUvdO	7	-	
N-	A	NAtrA	1	-		NAtlI	7	NAdzO	1	-	
	E	NEdzA	4	-		NEnrI	2	-		-	
	I	NImlA	2	NIrnE	3	-		-		-	
	O	NOrmA	2	-		-		-		NUmcU	5
	U	-		-		-		-			
P-	A	-		PAtcE	5	PArtI	27	PAskO	3	-	
	E	PEntA	2	-		PEtcI	2	PEnsO	6	PErnU	8
	I	-		-		-		PIfnO	1	PIskU	1
	O	POndA	7	-		POrlI	4	-		POnsU	7
	U	PUrdA	7	PUrfE	3	-		PUctO	7	PUntU	6
R-	A	*gRAsA	1	RAndE	1	*tRAtI	5	*bRAtO	4	RAtcU	1
	E	REtcA	1	-		REvrI	2	REnrO	5	*tRElU	1
	I	*bRIgA	3	*tRImE	2	RIlrI	3	-		*sRIsU	1
	O	*mROzA	3	-		*gROcI	1	-		*bROcU	2
	U	RUtmA	6	*bRUtE	1	RUlnI	7	-		*pRUtU	1
S-	A	SAnpA	13	SAnsE	4	SAntI	3	SAmtO	5	-	
	E	SEtfA	11	-		SEtcI	15	-		-	
	I	SItfA	23	-		SImcI	4	SIstO	23	SIltU	5
	O	SOndA	8	SOrmE	3	SOnlI	1	-		-	
	U	SUdnA	4	-		SUndI	7	SUnhO	6		
T-	A	TAknA	27	TArlE	1	*sTAlI	2	TArgO	4	TArmU	9
	E	-		-		TEdjI	1	-		TEtcU	1
	I	TIsrA	8	*sTIrE	1	-		TIdjO	1	TIfrU	3
	O	TOknA	6	TObmE	2	TOgrI	8	-		-	
	U	TUrkA	4	TUglE	2	*sTUlI	1	-		-	
V-	A	VAlnA	1	-		VAtlI	2	VAprO	2	VAlpU	2
	E	VEdmA	3	-		VEtcI	12	VEslO	11	VEndU	1
	I	VIzkA	6	-		(VIdrE)		-		VIdjU	2
	O	VOlsA	2	-		VOltI	3	-		-	
	U	-		-		-		-		-	
Z-	A	-		-		-		ZAvlO	1	-	
	E	-		-		-		-		-	
	I	-		-		-		-		-	
	O	-		-		-		-		-	
	U	-		-		-		-		-	

Number of Affixes: Number of Terms Covered

--A	--E	--I	--O	--U
60:166	20: 29	60:142	38:352	38: 59

TOTALS: 216: 1420

By Nat'ls: 925
By Unnat'ls: 279
Total: 1204
"Good" Unnaturals: *

CCV-ASSIGNMENTS

Shows the 36 Active Initial CC's
1204 Terms Covered; 77% by Naturals

180 Possible
123 Assigned
68% Used
Special Unnaturals: **

	-A	BLAda	3		DJAno	17		KRAku	10		-	
	-E	BLEka	8		DJEla	2		KREni	3		-	
BL-	-I	BLIcu	9	DJ-	DJIne	18	KR-	KRIdo	39	SR-	SRIte	19
	-O	BLOda	7		DJOri	9		KROli	5		*SO/Rdi	8
	-U	BLUdi	1		DJUdi	8		KRUma	6		*SU/Rna	3
	-A	BRAna	14		DRAni	2		*MA/Rka	6		STAdi	10
	-E	BREdi	2		DREti	6		MREnu	15		STEti	6
BR-	-I	BRIze	8	DR-	DRIki	8	MR-	-		ST-	STIse	10
	-O	BROko	5		-			*MO/Rdu	45		STOlo	22
	-U	BRUdi	6		*DU/Rzo	138		-			STUci	2
	-A	CKAno	6		DZAbi	2		PLAta	4		TCAro	6
	-E	CKEmo	21		DZEli	1		PLEci	9		TCEru	5
CK-	-I	-		DZ-	-		PL-	PLIzo	29	TC-	**TiTCI	13
	-O	CKOzu	21		DZOru	4		-			**ToTCO	5
	-U	-			-			PLUci	3		TCUre	17
	-A	CLAdo	6		FLAmi	1		PRAse	11		TRAdu	12
	-E	CLEsi	15		FLEti	2		**PaPRE	9		TREna	6
CL-	-I	CLIka	83	FL-	FLIdu	7	PR-	PRIre	15	TR-	TRIcu	6
	-O	CLOri	1		FLOfu	1		PROju	14		TROku	5
	-U	CLUva	7		(FLUro)			PRUci	4		TRUke	1
	-A	CMAlo	21		*FA/Rfu	3		SKAlu	7		-	
	-E	CMEni	6		FREna	13		-			TSEro	14
CM-	-I	CMIza	4	FR-	-		SK-	SKItu	5	TS-	TSIme	4
	-O	-			*FO/Rma	28		SKOri	2		-	
	-U	-			FRUta	3		*SU/Ksi	1		TSUfi	1
	-A	*CA/Nli	3		-			-			VLAko	1
	-E	-			-			-			-	
CN-	-I	CNIda	7	GL-	GLIda	2	SL-	SLIti	2	VL-	-	
	-O	-			-			SLOpu	1		-	
	-U	**CNiNU	6		-			SLUko	1		-	
	-A	-			GRAda	9		SMAno	3		-	
	-E	-			GREsa	12		-			VREti	4
CP-	-I	-		GR-	GRItu	6	SM-	SMIke	8	VR-	VRIci	4
	-O	-			GROda	21		-			-	
	-U	CPUla	7		GRUpa	30		SMUpi	1		-	
	-A	CRAno	1		-			*SA/Nca	4		-	
	-E	-			-			SNEku	1		-	
CR-	-I	CRIna	2	JM-	JMIte	6	SN-	SNIre	14	ZB-	-	
	-O	(CROmi)			-			(SNOla)			-	
	-U	-			-			-			ZBUma	8
	-A	-			KLAbu	3		SPAsi	16		-	
	-E	CTEki	3		KLEsi	9		SPEni	6		-	
CT-	-I	CTIfu	15	KL-	KLIri	8	SP-	SPIcu	7	ZV-	-	
	-O	-			KLOgu	2		SPOpa	4		ZVOto	10
	-U	CTUda	1		*KUt/La	7		SPUro	5		-	

Badjo	baj		bough
badlo	bad		bundle
bakso	bao		box
bakto	bak		bucket
balci	bac	bai	build
balma	bam		ball
balpi	bal		balance
banci	ban		bathe
banko	-		bank
banse	-		basket
barda	-		reward
barma	bar	baa	arm
basni	bas		base
batmi	bat		trade
batpi	bap		bottle
batra	-		butter
bekli	bel		bell
bekti	bek		object
bendu	ben	beu	band
berci	-		sheep
berti	ber	bei	carry
betcu	bet		bent
betpu bedpu	bed		bed
bidje	bie		edge
bilca	bic		military
bilti	bil	bii	beautiful
bisli	bis		ice
bisti pasko	pas	pak pao	past
bitsa	bit	bia	between
bivdu	biv	bid biu	behave
blabi	lab		white
blabo bulbi	bub	bui	bulb
blada	bla		blade
bleci blicu	bli		possible
bleka	ble		watch
blice bitce	-		whip
bloda	blo		hit
bludi	blu		blood
bongu	bon		bone
borku	bor		bow
botci	boi		boy
botcu botsu	bot	bou	boat
botni	-		button
bradi brudi	bru		brother
brana	bra		born
brano breba	reb		bread
brato	rao		ratio
bredi	bre		ready
breko	rek		brake
briga	rig	ria	brave
briku	rik		brick
brili	ril		bright
brize	bri		wind
brocu	roc	rou	brush
broda	rod		broken
broko	bro		break
brute	rue		breathe
bukcu	buk	buu	book
bulju	bul		boil
bunbo	bun		fool(ish)

buste	bus	bue	step
Cabro	cab	cao	burn
canse	can	cae	chance
capri	cap		print
carbo	car		carbon
cartu	cat		map
caslo	cas		whistle
cavle	cav		shovel
cefli	cef		chief
cenja	cen	cea	change
cidja	cid		awake
cidza cedzu	ced		shadow
ciktu	cik	ciu	equal
cimra	cim		summer
cinta	cin		infant
cirna	cir	cia	learn
cirzi cersi	cei		chair
cirzo	cio		scissors
ckano	cka		kind
ckela	kea		school
ckemi kemdi	kem		chemical
ckemo	cke		time
ckozu	cko		cause
clado	cla		loud
claso clesi	cle		without
clidu	-		slide
clife	lif		leaf
clika	cli		like
clina	lin	lia	line
clivi	liv	lii	live
clivu cluva	clu		love
cmalo	cma		small
cmeni	cme		money
cmiza	cmi		amuse
cnida	cni		need
cninu	cnu		new
cnire snire	sni		near
colku	col		silk
colri clori	clo		chlorine
comtu	com	cot cou	ashamed
condi	con		deep
condu hanco	han		hand
corta	cor	coa	short
crano	cra		smile
crina	cri		rain
cteki	cte		tax
ctifu	cti		stuff
cundo	cuo		window
cupri	cup		copper
cutci	cuc		shoe
cutri	cut	cui	water
cutse	cus	cue	say
Dakli	dak		probable
damni	dam		down
dampa	-		pump
danci	dan	dai	plan
danri	-		ordinary
dante	dat		tooth

danza		daa		desire
darli		dar		far
darto		dao		door
denli		del	dei	day
denro		den		danger(ous)
dertu		deu		dirt
detri	detra	der	dea	daughter
dilri		-		represent
dipri		dip		precious
dirci	dirco	dir	dio	direction
dislu		diu		discuss
disri		dii		decide
ditca		dia		teach
ditcu	citlu	cil		detail
ditka		dit		bite
ditlu	targo	tag	tao	argue
djadi	djudi	dju		judge
djale		jal		ring
djano		dja		know
djela		dje	jel	healthy
djeta		jet		owe
djeto	dedjo	ded	dej	finger
djimi	jmite	jmi		meet
djine		dji		join(t)
djino	vinjo	vin		wine
djipo		jip	jio	important
djiri	hijra	hia		here
djitu		jit	jiu	tight
djora	jorta	jor	joa	hour
djori		djo		member
djoso		jos		sew
djula		jul		jewel
djupo		jup	juo	support
donsu		don	dou	give
dorja		dor		war
dotra		dot		winter
draka		rak		dark
drara		-		drawer
dreti		dre		correct
driki		dri		remember
dumni	humni	hum	hun	human
durna		dun	dua	adorn
durzo		dru	dur	do
dustu		dus		dust
duvra	duvri	duv	dui	discover
dzabi		dza		real/exist
dzaso		zas		soap
dzeli		dze		jelly
dzoru		dzo		walk
dzozo	hozda	-		hose
Fagro		fag		fire
falba		fab	faa	fail
falji		fal	fai	false
famji		fam		family
fanpo	pozfa	poz		oppose
fanra		fan		farm
fanri	fanve	fav		reverse
farfu		fra	far	father
fasli		-		face
fasru		fas		easy
fekti	purfe	puf	pue	perfect
fekto		fek	feo	fact
femdi		fem		female
ferci		fei		affair
ferlu	felda	fel	fed fea	fall
ferno	fernu	fen	feu	iron
ferti		fet		fertile
festi		fes		waste
fibru		fib		weak
fikco		fik		fiction
fildi		fii		field
filmo		fil	fio	feel
firpa		fir	fia	afraid
fitpi		fit	fip	foot
fizdi		fiz	fid	physical
flami		fla		flame
fleti		fle		fly
flofu		flo		float
flora		lor	loa	flower
foldi		fod		fold
folma		fol	foa	full
forli	fotli	fot	foi	strong
forma		fro	fom	form
fosli		fos		force
fragu		rag		fog
frama		ram		frame
fremi		rem		friend
frena		fre		front
frezi		rez		free
fruta		fru		fruit
fulri		-		rich
fumna		fum	fua	woman
fundi		-		like/fond
futci		fut	fuc fui	future
Gacpi	hapci	hap	hai	happy
gancu		gac	gau	win
gandi		gad		god
ganli		gal		organize
ganta		gan	gaa	high
gardi		-		garden
garko	harko	har	hao	shelter
garni		gar	gai	rule
garti		gat		grateful
gasno		gas		anus
gasti		-		steel
genza		gen	gea	again
getsi		get	gei	get
ginru		gin		root
girsa	gresa	gre		grease
glida		gli		guide
gliso		-		glass
gokru		gok		hook
gomni		gom		sticky
gorma	hatro	hat		hot
gotca		goa		goat
gotri		gor		industry
gotso		got	gos goo	go
grada		gra		great

grani drani dra dry
grasa raa grass
grato - cake
gritu gri sing
groci roi angry
groda gro big
gruni rulni rul rui rule
grunu run grain
grupa gru group
gudbi gud gub good
gunti gun gui country
gusti kusti - costly
gusto gus gut flavor
gutra gur strange

Jaglo jag jao angle
jalti jai product
janro jar narrow
janto jan hunt
jokla jok clock
jugra jug jua grab
junti jun young
jupni jui opine
jurna - earn

Kakto kak kao act
kalpi kopca kop copy
kamda kad fight
kamfu komfu kof kou comfortable
kamla kam kaa come
kamra kar camera
kamti - committee
kamtu kumtu kum kuu common
kance kac kae conscious
kanci kunci kuc kui relation
kangu kau dog
kanli canli cna cai quantity
kanlo klogu klo close
kanmo kan able
kanpi - compete
kanra - cane
kanse perti - concern
kanta celna cel shelf
kante konte kon count
kanti - bill
kanto troli tol control
kanvi vizka viz vik via see
kapli kal complete
kapma - hat
kapni kap open
karda - card
karku - crack
karsa kas across
kasfa kaf punish
kasni - cow
katca - watch
katli kat kai quality
katma - cat
kecri kec kei sad
kerfa herfa her hair
kerju kej keu care
kerti ker air
ketli ket kettle
kicmu kic kiu doctor
kinci kin kii companion
kinku kik sharp
klabu kla cloth
klada - cloud
klesi kle class
klini - clean
klipu - keep
kliri kli clear
kokfa kok cook
koldu holdu hol hou hole
kolro kol koo color
komcu kom comb
kompi - company
konsu kos consul
korci korce koc koe cord
korji koj koi command(er)
korka - cork
korma horma hor hoa horse
korti kor body
korva kov koa curve
kraju raj scratch
kraku kra cry
krali kroli kro current
kreni kre ray
krido kri believe
krinu - nut
kruma kru room
kubra kub wide
kukra kuk fast
kunti kenti ken question
kupta kup cup
kurfa kur square
kurni kun warn
kusfa hasfa has haf haa house
kusmo kus kuo custom
kuspo - spread
kutla klu cut
kuvga kuv kua cover

Lakse lak wax
laksu sluko slu lock
laldo lal old
landi lan lai land
larte lar lae art
lasti las elastic
ledri led lightning
ledzo lez left
lelpi lel lep level
lengu len language
lenki lek electric
lenri nenri nen nei in
lenze lenzo leo lens
lerci ler letter
lesta les east
letci let lei let
letra lea character

lidro	hidro	hid		hydrogen
likro		lio		liquor
likta		lik		week
likti	flidu	fli		liquid
lilfa		lil		legal
lilpa	porli	por	poi	power(ful)
limji		lim	lij	limit
linco		lic		thin
lista		lis		list
litla		lit		light
litnu		liu		hold
litri	cisti	cit		history
lodji		lod		logic(al)
lokti		loi		local
lunli		lul		wool
lusta		lus	lut	west
Madzo		mad	mao	made
malbi		mal		sick
malna		man		milk
mandi	mendi	men		male
mandu	dupma	dup		deceive
manko		mak		mouth
manta	monca	mon	moa	mountain
marka		mra	mar	mark
matci		mai		machine
matma		mat	mam maa	mother
menki		mek		eye
mensa		mea		month
merji		mer		marry
merli		mel	mei	measure
metli		met		metal
metlo	cetlo	cet		wet
metri	sorme	som	soe	sister
midju		mid	mij	middle
miksa		mis		mix
mildo		-		mild
milfa		mia		meal
minku		miu		ore
mipli		mip	mii	example
mitro		mit	mio	meat
modvi		mov	moi	motive
monza		moz		morning
mordu		mro		more
morto		mor		dead
motci		moc		motor
mrenu		mre		man
mroza		roz	roa	hammer
mubre		mub		wood
murki		muk		monkey
mursi		mur		sea
muslo		mus		muscle
muvdo		muv	muo	move
muzgi		muz	mui	music
Nadzo		naz	nao	now
najda		naj		knife
naldi		nal		nail
namci		nam		name
nanta	nanda	-		knot
nardu		nad		difficult
narmi		-		army
narti		-		apart
natli		nat	nai	night
natra		nar	naa	nature
natri	nadri	-		sodium
nedza		ned	nea	next
negda		neg		egg
nerbi		neb		necessary
nervi		nev		nerve
nesta		nes		honest
nigro		nig		black
nikri		nik		cheese
nilca		-		below
nimla		nim	nia	animal
nirda		-		bird
nirli		nil		girl
nirne		nin	nie	year
nitci		nit		neat
nitru	fatru	fat	fau	trouble
norla	nordi	nod		north
norma		nom	noa	average
norsa	notbi	not		other
nreti	netre	net		net
nrile	ridle	rid		read
numcu		num	nuu	number
nuzvo		nuz		news
Packo	packe	-		pocket
padzi		paz		pad
pafko		-		dig
palna	nable	nab		problem
panba		pan		pan
panta	herba	heb		plant
pante	patpe	pat		pot
papre		pre		paper
parte	parti	par	pai	part
pasti		-		paste
pazda		pad		wait
penbi		-		pen
pendi		ped		hang
penso		pen	peo	think
penti	petci	pec	pei	pay
pento	penta	pet	pea	point(ed)
pernu		per	peu	person
petri		-		distribute
pidra	hompi	hom		drink
pifno		pif	pio	frequent
pilno		pil		plain
pinca		pic		urine
pinda		pid		pin
pinti		pin		paint
pisku		piu		piece
pismi		pis		peace
pismu	smupi	smu		smooth
plado		lad		plow
plata		pla		plate
plici	pleci	ple		play
plizo		pli		use
pluci		plu		please

pluma	lum lua		feather
podju proju	pro		produce
poldi	pod		nation
ponsu	pon pou		own
porju	poj		pig
posta	pot		post(al)
potri hutri	hut hui		destroy
pozbu bufpo	buf		opposite
prali	ral		profit
prano	-		run
prase	pra		continue
prati	rat		price
preni	-		prisoner
prire	pri		behind
prizi	riz		private
proza	-		prose
pruci	pru		test
prusa	rus		approve
prutu	ruu		protest
publi	pub		public
pubmi pubmu	pum		lead
puctu pucto	puc puo		push
pudru	pud		powder
pulso	pul		impelled
punfo	-		pure
puntu	pun puu		pain(ful)
purda	pur pua		word
Randi rande	rad rae		round
rando fando	fad fao		end
ranjo	-		range
ranta	ran		rotten
rasto	ras		brass
ratci patce	pac pae		device
ratcu	rau		rat
raznu	raz		reason
redro	red		
renro	ren reo		throw
resfu	res ref		dress
resra	-		restaurant
resta zbuma	zbu		explode
resto	ret		recline/rest
retca	rec rea		different
retpi	rep		answer
revri	rev rei		dream
ridji lidji	lid		religious
rilri	rii		regular
rinta	rin		rhythm(ic)
rirda	rir		record
rispa daspa	das dap		responsible
rispe	ris		respect
ritco	ric		right
rodja	roj		grow
rodlu	-		road
rofsu	rof ros		rough
rolgu	rol		roll
rutma	rut rua		route
Sakli	sak		sack
saldi	sal		solid
salfa	saf		sail
samto	sam sao		same
sanca	sna		sand
sange	-		suggest
sanpa	san saa		sign
sanse	sas sae		sense
santi	sai		silent
sapla	sap		simple
sarni	sar		sour
satci	sac		start
satro	sat		rub
sedji selji	sel		self
sekci	sec		sex(ual)
selba helba	hel		help
sensi	ses		science
setci	set sei		set
setco sekta	sek		insect
setfa	sea		put
sidza	siz		seed
siltu	sil siu		shake
simba	sib		lion
simci	sim sii		seem
sinma	sin		cinema
sinta snatu	-		tin
sirna	sir		certain
sitci	sic		city
sitfa	sit sif sia		place
sitmo sisto	sis sio		system
skalu	ska		scale
skapi	-		skin
skiti begco	beg beo		request
skitu	ski		sit
skizo	kiz		ski
skori	sko		screw
sliti	sli		sweet
slopu	slo		steep
smano	sma		smoke
smike	smi		secret
smina	min		mind
sneku	sne		neck
snice	nic		snow
socli	sol		social
solda	sod		soldier
solte	sot		salt
sonda	son soa		sound
sonli	soi		sleep
sonta sunho	sun suo		son
sordi	sro		store
sorlu sorgu	sog		ear
spada ponda	poa		respond
spali	pal		side
spana hospa	hos		hospital
spasi	spa		space
spebi	peb		special
spena penja	pej		sponge
speni	spe		experience
spicu	spi		spirit
spila cpula	cpu		pull
spopa	spo		hope
sporu	-		spring

spuro	spu		skill(ed)
sputa	put		spoon
sputu	spetu	-	spit
srisu	riu		serious
srite	sri		write
stadi	sta		stage
stali	tai		stand
stana	-		station
stane	staga	-	stem
stari	-		surprise
stire	tir	tie	stairs
stise	sti		stop
stisi	steti	ste	sentence
stolo	sto		stay
stuci	stu		story
studa	ctuda	ctu	feces
stuli	tui		adjust
sucmi	-		swim
sudna	sua		sudden
suksi	sku		succeed
sulba	sul		swelling
sumji	sum		sum
sundi	sud	sui	send
supta	sup		soup
surla	surdi	sur	south
surna	sru		injure
surva	suv		serve
sutme	sut		smell
Takna	tak	taa	talk
tarci	tar		star
tardu	hardu	had	hard
tarle	tal	tae	tired
tarmo	harmo	ham	harmony
tarmu	tam	tau	weapon
tcaku	cak		shock
tcale	langa	lag laa	long
tcali	cal		wall
tcari	karti	-	cart
tcaro	tca		car
tcela	-		wing
tcena	-		chain
tceru	tce		through
tcori	-		authority
tcura	curca	cur	safe
tcure	tcu		picture
tedji	ted	tei	attend
tenri	ten		increase
tepli	tep		church
terla	tel		terrestial
testi	-		gonad
tetcu	tec	teu	stretch
tetri	tet		weather
tidjo	tid	tio	heavy
tifru	tiu		offer
tinmo	tin		ink
tirca	tic		wire
tirku	hirti	hir	hear
tisra	tia		selec
titci	tci		eat
tobme	tob	toe	table
togri	tog	toi	agree
tokna	toa		take
tokri	-		chalk
tomti	tomki	tok	automatic
torni	ton		twist
tosku	hedto	hed	head
totco	tco		touch
totnu	tot		thick
tovru	tov		over
traci	rac		travel
tradu	tra		true
trana	tan		rotate
trani	trali	-	tray
trati	rai		try
treci	-		interesting
trelu	rel	reu	rail
trena	tre		train
tricu	tri		tree
trida	-		street
trili	til		attract(ive)
trime	rim	rie	tool
troku	tro		rock
trufa	ruf		roof
truke	tru		structure
tsero	tse		error
tsime	tsi		crime
tsufi	tsu		enough
tubli	tub		tube
tugle	tul	tue	leg
turka	tur	tua	work
Valda	vad		develop
valna	van	vaa	violent
valpu	vap	vau	wave
vapra	vapro	vao	gas
vatli	val	vai	value
vedji	mutce	muc mut mue	much/very
vedma	ved	vea	sell
vendu	ven	veu	poison
verti	vreti	vre	vertical
veslo	veo		vessel
vetci	vet	vei	happen
vetfa	-		invent
vidju	vij	viu	view
vidre	vid		idea
virsa	vis		poetry
virta	vir		ad
visra	-		viscera
vlako	vla		lake
volsa	vol	voa	voice
volti	vot	voi	jump
vrici	vri		river
Zavlo	zav	zao	bad
zavno	zan		oven
zvoto	zvo		out

New	Old	English
Bedpu	betpu	bed
begco	skiti	request
bitce	blice	whip
blicu	bleci	possible
botsu	botcu	boat
breba	brano	bread
brudi	bradi	brother
bufpo	pozbu	opposite
bulbi	blabo	bulb
Canli	kanli	quantity
cedzu	cidza	shadow
celna	kanta	shelf
cersi	cirzi	chair
cetlo	metlo	wet
cisti	litri	history
citlu	ditcu	detail
clesi	claso	without
clori	colri	chlorine
cluva	clivu	love
cpula	spila	pull
ctuda	studa	feces
curca	tcura	safe
Daspa	rispa	responsible
dedjo	djeto	finger
detra	detri	daughter
dirco	dirci	direction
djudi	djadi	judge
drani	grani	dry
dupma	mandu	deceive
duvri	duvra	discover
Fando	rando	end
fanve	fanri	reverse
fatru	nitru	trouble
felda	ferlu	fall
fernu	ferno	iron
flidu	likti	liquid
fotli	forli	strong
Gresa	girsa	grease
Hanco	condu	hand
hapci	gacpi	happy
hardu	tardu	hard
harko	garko	shelter
harmo	tarmo	harmony
hasfa	kusfa	house
hatro	gorma	hot
hedto	tosku	head
helba	selba	help
herba	panta	plant
herfa	kerfa	hair
hidro	lidro	hydrogen
hijra	djiri	here
hirti	tirku	hear
holdu	koldu	hole
hompi	pidra	drink
horma	korma	horse
hospa	spana	hospital
hozda	dzozo	hose
humni	dumni	human
hutri	potri	destroy
Jmite	djimi	meet
jorta	djora	hour
Karti	tcari	cart
kemdi	ckemi	chemical
kenti	kunti	question
klogu	kanlo	close
komfu	kamfu	comfortable
konte	kante	count
kopca	kalpi	copy
korce	korci	cord
kroli	krali	current
kumtu	kamtu	common
kunci	kanci	relation
Langa	tcale	long
lenzo	lenze	lens
lidji	ridji	religious
Mendi	mandi	male
monca	manta	mountain
mutce	vedji	much/very
Nable	palna	problem
nadri	natri	sodium
nanda	nanta	knot
nenri	lenri	in
netre	nreti	net
nordi	norla	north
notbi	norsa	other
Packe	packo	pocket
parti	parte	part
pasko	bisti	past
patce	ratci	device
patpe	pante	pot
penja	spena	sponge
penta	pento	point(ed)
perti	kanse	concern
petci	penti	pay
pleci	plici	play
ponda	spada	respond
porli	lilpa	power(ful)
pozfa	fanpo	oppose
proju	podju	produce
pubmu	pubmi	lead
pucto	puctu	push
purfe	fekti	perfect
Rande	randi	round
ridle	nrile	read

Sekta	setco	insect
selji	sedji	self
sisto	sitmo	system
sluko	laksu	lock
smupi	pismu	smooth
snatu	sinta	tin
snire	cnire	near
sorgu	sorlu	ear
sorme	metri	sister
spetu	sputu	spit
staga	stane	stem
steti	stisi	sentence
sunho	sonta	son
surdi	surla	south
Targo	ditlu	argue
tomki	tomti	automatic
trali	trani	tray
troli	kanto	control
Vapro	vapra	gas
vinjo	djino	wine
vizka	kanvi	see
vreti	verti	vertical
Zbuma	resta	explode

AFFIX USAGE & PRIMITIVE POWER

Last Update: 2 Aug 82

In this listing the primitives are arranged in the order of their "power": the number of terms in the L4 pool of complexes to which each primitive contributes short-affixes. For example, madzo, which makes 277 such contributions, has the greatest power. The primitives with their affixes are arranged in groups of descending power. The number of times each affix is used under the current strategy for remaking the L4 CPXs is also shown. If that strategy were to be changed--if the vowel-rich strategy discussed elsewhere in this Notebook were, for example, to be adopted--the affix usage data would be slightly different.

Usage-data on the three main types of affixes are shown in separate columns in the tables, and summary statistics are given for each type at the end of each group. For all groups but the first, accumulative statistics are given on a final line which includes that group with all groups of higher power.

No. of Terms	Old Prim	As Remade	Affix Types & Coverage CCV	n	CVC/C	n/n	CVV	n	Coverage Loss	English Key-Word	
277	madzo				mad	14	mao	263	-	made	
156	durzo		dru	138	dur	18			-	do	
83	clika		cli	83					-	like	
69	folma				fol	3	foa	66	-	full	
63	cenja				cen	6	cea	57	-	change/become	
54	gotso				got/s	11/3	goo	40	-	go	
50	forma		fro	28	fom	22			-	form	
45	mordu		mro	45					-	more	
44	cutse				cus	2	cue	42	-	say	
43	kakto				kak	18	kao	25	-	act	
40	takna				tak	13	taa	27	-	talk	
39	krido		kri	39					-	believe	
37	parte	parti			par	10	pai	27	-	part	
34	donsu				don	2	dou	32	-	give	
32	sitfa				sit/f	6/3	sia	23	-	place	
30	grupa		gru	30					-	group	
29	plizo		pli	29					-	use	
28	sitmo	sisto			sis	5	sio	23	-	system	
											Terms
28+	18 wds	2	7	392	15	136	11	625	0	33 afs	1153
	2.6%	11%	56		9.7		56.8		0%	4%	25%
Final Total:	702 wds									815 afs	4421

The final totals show that, by the end of the list, 702 primitives will be assigned 815 affixes which will, collectively, cover 4421 terms. To this must be added the 236 terms which will remain uncovered, yielding 4421 + 236 or 4657 as the size of the total pool of terms. It is of this total, for example, that the 1153 terms covered by the 33 affixes in this first group constitute 25%. Thus, 2.6% of the words and 4% of the affixes cover 25% of the terms.

Note how important the CVV affixes are in this group. They cover more than half (625/1153 or 54%) of all the terms its affixes cover. The average coverage of the 11 CVVs is 56.8 terms, slightly higher than the 7 CCVs at 56 terms. Even setting aside the extreme cases mao and dru, the average contribution of the

other CVVs, at 36.2 terms, approaches that of the other CCVs, namely 42.3 terms, and it is four times as great as the mean coverage of the 15 CVCs at 9.1 terms per affix. Thus, the CVV-form is very important at the high end of the power scale. But this importance will diminish as the power of the prims diminishes.

The remaking rate--here 11%--is also an interesting statistic. This is the lowest value it will have in the power-listing. What this means is that these most powerful words tend not to be packed, and that even when they are packed it is the weaker and less common words that compete with them that were most profitably changed.

In the next group, the first coverage losses appear. Whether the uncovered terms involved are final ('f') or non-final ('n') is noted in the entry. The accumulating records of the two types of losses are then kept separately.

27	clivi				liv	21	lii	6	-	live
	ganta				gan	26	gaa	1	-	high
	garni				gar	23	gai	4	-	rule
25	katli				kat	8	kai	17	-	quality
24	kanli	canli	cna	3			cai	21	-	quantity
	sensi				ses	24			(21f)	science
22	sanpa				san	9	saa	13	-	sign
	stolo		sto	22					-	stay
21	ckemo		cke	21			(keo)		-	time
	ckozu		cko	21					-	cause
	cmalo		cma	21					-	small
	groda		gro	21					-	big
	ponsu				pon	14	pou	7	-	own
20	gudbi				gud/b	16/2			2f	good
	kusfa	hasfa			has/f	7/2	haa	11	-	house
	sedji	selji			sel	20			-	self
19	lenri	nenri			nen	17	nei	2	-	in
	letci				let	3	lei	16	-	let
	pernu				per	11	peu	8	-	person
	srite		sri	19					-	write

27-19:	20 wds	4	7	128	15	203	11	106	0 2	33 afs	437
19+:	38	6	14	520	30	339	22	731	0 2	66	1590
	5.4%	20%	18.3		13.5		9.6		0.5%	8.1%	34%

With the addition of this group to the first, more than a third of the pool is now covered. But note that the CVVs are diminishing in importance, having an average coverage of 9.6 terms in this group. The CVCs are increasing in relative importance, yielding 13.5 terms each, and the CCVs are now distinctly the most useful, being used 18.3 times per affix. cna is a special case. Unnatural as cna is, the tastier cai is used whereever it can be, and cna used only where cai can't. can and cal are not available, being more usefully assigned to canse (of power 10) and tcali (a 4).

Our first coverage loss occurs in this group: 2 gudbi-final words (bilgudbi and mucgudbi) are not covered. gui is much more usefully assigned to gunti (a 15)

and there is no possible CCV. Notice that the 21 sensi-final words are not reckoned as coverage losses. We actually want the final term to be unreduced in these words. The lesta- and lusta-final words (e.g., surlesta) are also desirably long, and so not counted in the pool of terms.

Note that the remaking rate has doubled. It has gone from 11% to 20% It will continue to rise as the primitives decrease in power until it peaks at around 30% in the neighborhood of the 8's. Then it will go down again as occasions for solving tuning problems by remaking words diminish.

18	djine		dji	18					-	join(t)
	futci				fut/c	10/2	fui	6	-	future
	kapli				kal	17			1f	complete
	kuvga				kuv	4	kua	14	-	cover
									-	
17	dirci	dirco			dir	5	dio	12	-	direction
	djano		dja	17					-	know
	fanri	fanve			fav	17	(fae)		-	reverse
	kanci	kunci			kuc	1	kui	16	-	relation
	purda				pur	10	pua	7	-	word
	setci				set	2	sei	15	-	set
	tcure		tcu	17					-	picture
	zavlo				zav	16	zao	1	-	bad
									-	
16	bisti	pasko			pas/k	10/3	pao	3	-	past
	kanvi	vizka			viz/k	8/2	via	6	-	see
	muvdo				muv	9	muo	7	-	move
	spasi		spa	16					-	space
	tarmu				tam	7	tau	9	-	weapon
	vetci				vet	4	vei	12	-	happen
									-	
15	claso	clesi	cle	15					-	without
	ctifu		cti	15					-	stuff
	gunti				gun	10	gui	5	-	country
	kamla				kam	5	kaa	10	-	come
	mrenu		mre	15					-	man
	prire		pri	15					-	behind
	sonda				son	7	soa	8	-	sound

18-15	25 wds	6	8	128	20	149	15	131	0 1	43 afs		408
15+	63	12	22	648	50	488	37	862	0 3	109		1998
	9.0%	24%	16		7.5		8.7		0.2%	13%		43%

We now have about 9% of the prims and 13% of all the affixes that will be assigned; and we have already covered 43% of the pool of terms. The CVCs and CCVs are of about equal importance in this group, covering 7.5 and 8.7 terms each respectively. The CCVs are still most powerful, covering 16 terms each.

Again, a small coverage loss has been sustained: prakapli with its unreduced final term is only partly covered. (kai has already been given to the more powerful katli, a 25.) Note that the remaking rate at 24% is still climbing.

14	brana		bra	14					-	born
	cnire	snire	sni	14					-	near
	filmo				fil	10	fio	4	-	feel

	gruni	rulni			rul	7	rui	7	-	rule
	lilfa				lil	13			1f	legal
	podju	proju	pro	14					-	produce
	tsero		tse	14					-	error
									-	
13	berti				ber	3	bei	10	-	carry
	bivdu				biv/d	6/1	biu	6	-	behave
	frena		fre	13					-	front
	jaglo				jag	8	jao	5	-	angle
	kinci				kin	12	kii	1	-	companion
	norsa	notbi			not	13			-	other
	numcu				num	8	nuu	5	-	number
	titci		tci	13					-	eat
	turka				tur	9	tua	4	-	work
	vedji	mutce			muc/t	10/2	mue	1	-	much/very

14-13	17 wds	5	6	82	13	102	9	43	0 1	28 afs	227
13+	80	17	28	730	63	590	46	905	0 4	137	2225
	11%	29%	13.7		7.8		4.8		0.4%	17%	48%

The CVCs are now more powerful, at 7.8 terms each, than the CVVs at 4.8 terms. The CCVs are still most powerful at 13.7 terms each. This pattern will be maintained, with minor variations, through the rest of the list. Again a CPX with an unreduced final term (blililfa) is encountered. (lia more fruitfully belongs to clina, a 12.) Total coverage-loss now stands at 4 terms, all final ones. The rate of loss--now 0.4%--is still a mere trickle. This pattern, too, will be maintained until we near the end of the list. The remaking rate has now peaked at 29%. But this is one of two peaks in the remaking rate; the other and broader one will not come until the 8's.

12	clina				lin	8	lia	4	-	line
	cutri				cut	9	cui	3	-	water
	denli				del	3	dei	9	-	day
	fumna				fum	2	fua	10	-	woman
	gacpi	hapci			hap	10	hai	2	-	happy
	girsa	gresa	gre	12					-	grease
	landi				lan	5	lai	7	-	land
	merli				mel	3	mei	9	-	measure
	morto				mor	12			-	dead
	penso				pen	6	peo	6	-	think
	satci				sac	11			1f	start
	skiti	begco			beg	2	beo	10	-	request
	tradu		tra	12					-	true
									-	
11	forli	fotli			fot	10	foi	1	-	strong
	merji				mer	10			1f	marry
	poldi				pol/d	10/1			-	nation
	prase		pra	11					-	continue
	puntu				pun	5	puu	6	-	pain(ful)
	setfa						sea	11	-	put
	socli				sol	11			-	social
	veslo						veo	11	-	vessel

12-11	21 wds	4	3	35	17	118	13	89	0 2	33 afs	242
11+	101	21	31	765	80	708	59	994	0 6	170	2467
	14%	19%	11.7		6.9		6.8		0.8%	21%	53%

We've passed the 50%-coverage mark with just 14% of the prims and 21% of the affixes. The locally increased utility of the CVVs will prove to be temporary. There is a continuing trickle of coverage losses, but both cases still involve only final terms. (sai will do more work for santi, a 4; and mei is already working strongly for merli, a 12.) These are, of course, the "tastier" of the two kinds of coverage losses, non-final ones always involving the not-very-tasty hyphens.

10	canse				can	8	cae	2	-	chance
	corta				cor	9	coa	1	-	short
	danci				dan	8	dai	2	-	plan
	detri	detra			der	3	dea	7	-	daughter
	farfu		fra	3	far	7			-	father
	korti				kor	9			1f	body
	kraku		kra	10					-	cry
	larte				lar	5	lae	5	-	art
	lilpa	porli			por	6	poi	4	-	power(ful)
	namci				nam	7			3f	name
	natli				nat	3	nai	7	-	night
	nimla				nim	8	nia	2	-	animal
	rando	fando			fad	7	fao	3	-	end
	sanse				sas	6	sae	4	-	sense
	stadi		sta	10					-	stage
	stise		sti	10					-	stop
	tcale	langa			lag	9	laa	1	-	long
	togri				tog	2	toi	8	-	agree
	vatli				val	8	vai	2	-	value
	zvoto		zvo	10					-	out

10	20 wds	4	5	43	16	105	13	48	0	4	34 afs		196
10+	121	25	36	808	96	813	72	1042	0	10	204		2663
	17%	20%	8.6		6.6		3.7		2.0%		25%		57%

The trickle of coverage losses has increased. But these latest 4 are still of the same type: like the others so far, they leave only unreduced final terms. This is surely the most tolerable kind of coverage loss, being hyphen-free.

9	bleci	blicu	bli	9					-	possible
	condu	hanco			han	9			-	hand
	djori		djo	9					-	member
	duvra	duvri			duv	2	dui	7	-	discover
	grada		gra	9					-	great
	kapni				kap	7			2f	open
	klesi		kle	9					-	class
	kolro				kol	4	koo	5	-	color
	korji				koj	2	koi	7	-	command(er)
	kukra				kuk	9			-	fast
	matci						mai	9	-	machine
	matma				mat/m	5/2	maa	2	-	mother
	midju				mid/j	8/1			-	middle
	nedza				ned	5	nea	4	-	next
	papre		pre	9					-	paper
	pento	penta			pet	7	pea	2	-	point(ed)
	plici	pleci	ple	9					-	play

	renro				ren	4	reo	5	-		throw
	resfu				res/f	6/3			-		dress
	sonta	sunho			sun	3	suo	6	-		son
	sundi				sud	2	sui	7	-		send

9	21 wds	6	6	54	17	79	10	54	0	2	33 afs			187
9+	142	31	42	862	113	892	82	1096	0	12	237			2850
	20%	29%	9		4.6		5.4		1.1%		29%			61%

The loss-rate has diminished a little; and it is still of the same favorable type. The remaking rate is 29% again. It will go no higher. We are evidently in the region of closely packed words. Moreover, the words now being remade are often of secondary importance. They were the ones that were preferentially remade when competing with a more powerful word; djadi with djano for dja, for example.

8	bilti				bil	4	bii	4	-		beautiful
	bleka		ble	8					-		watch
	botcu	botsu			bot	3	bou	5	-		boat
	brize		bri	8					-		wind
	djadi	djudi	dju	8					-		judge
	dorja				dor	8			-		war
	driki		dri	8					-		remember
	dumni	humni			hum/n	4/3				1f	human
	femdi				fem	8			-		female
	kliri		kli	8					-		clear
	mandi	mendi			men	8			-		male
	marka		mra	6	mar	2			-		mark
	nadzo				naz	7	nao	1	-		now
	prali				ral	7				1f	profit
	puctu	pucto			puc	1	puo	7	-		push
	resta	zbuma	zbu	8					-		explode
	ridji	lidji			lid	8			-		religious
	rodja				roj	7				1f	grow
	rutma				rut	2	rua	6	-		route
	saldi				sal	8			-		solid
	samto				sam	3	sao	5	-		same
	smike		smi	8					-		secret
	sordi		sro	8					-		store
	tisra						tia	8	-		select

---

8 :	24 wds	7	9	70	16	83	7	36	0	3	32 afs			189
8+:	166	38	51	932	129	975	89	1132	0	15	269			3039
	24%	29%	7.8		5.2		5.1		1.6%		33%			65%

Now with just one-quarter of the prims employed we have gained nearly two-thirds of the coverage. There is still the same sort of loss-trickle. Remaking is holding at its peak rate of 29%.

7	bilca				bic	7			-		military
	bitsa				bit	3	bia	4	-		between
	bloda		blo	7					-		hit
	clivu	cluva	clu	7					-		love