3-14-1989

A Computer Rhyming Dictionary/Research Tool

Melvin J. Luthy

Robert Stevens

Follow this and additional works at: http://scholarsarchive.byu.edu/dlls

A COMPUTER RHYMING DICTIONARY/RESEARCH TOOL

Melvin J. Luthy
Brigham Young University
Robert Stevens
WordPerfect Corporation

Recognizing the Need

For years poets, writers, linguists, scrabble players,
crossword puzzle addicts, language materials developers,
and other language lovers have had frequent occasion to
search for words having particular sounds or sound
combinations. The poet may search for a word beginning
with the right consonant to complement an alliterative
line, or a word with particular vowels to complete the
assonance in a line, or perhaps he may search for the right
final syllable for a perfect or identical rhyme. The
language materials developer may need to collect examples
of particular consonant clusters or vowel combinations to
help foreign students practice English phonological
patterns. The linguist may be interested in statistical
data such as frequency of occurrence of particular sounds,
or occurrence of particular phonotactic patterns, for which
English orthography is inadequate. Even "Dear Abby" needed
such information to respond to a writer who asked whether
there were more than two words (angry and hungry) that
ended in the sounds "gry." Abby could not think of others,
but offered an anecdote about a lady who approached George
Bernard Shaw with the observation that there was only one
word in English beginning with the letters su that had the
sound [šU]. That word was sugar. George Bernard Shaw
paused for a moment and replied, "Are you sure?"

Each of these circumstances illustrates the need for a way
to retrieve words from the English lexicon in a way more
rapid than introspection. To meet this need, rhyming
dictionaries have been published, but they are generally
limited to word-final rhymes, and are relatively time-
consuming. Of course the ideal solution would be a computer
program that would enable a user to select any sound(s) in
any word position(s), and let modern technology retrieve the
words with the corresponding sounds, be they vowels for end
rhyme or assonance, or consonants for alliteration or
consonance. Such a program should at least allow for the
use of key words as models, the selected sounds of which
would be matched by the computer in selecting corresponding
words. Ideally, it should also allow the user to select
sounds independent of any word, and thereby retrieve words
containing those sounds in the phonotactic positions
specified by the user. Finally, it would be a bonus to have
such a program interfaced with a thesaurus, so that one may
search for a word within a given semantic range containing a

particular sound or sound combination. Such a feature would help those afflicted with the "tip of the tongue" phenomenon.

Developing a Solution: Data Base

These are the goals of the current project which we are presenting today. In order to accomplish such a task, two things must be in place: First of all, there must be an accurate, consistent data base; and second, there must be an adequate computer program to retrieve the desired data. Melvin Luthy has been working on the data base, and Robert Stevens has been working on the computer program. At this point, we have 62,500 words in the base written in phonetic transcription with syllabification marked. The transcription system is modified IPA (International Phonetic Alphabet) consisting of the following symbols:

[p,b,t,d,k,g,f,v,θ,ð,s,z,š,ž,h,č,ǰ,m,n,ŋ, l,r,w,j, i,I,e,E,æ,ə,u,U,o,ɔ,ɚ,ai,aiɚ,au,auɚ,aɚ,ɔɚ,ɔi,Uɚ,Iɚ,Eɚ,æɚ]

Syllable boundaries are marked with a hyphen. Only segmental sounds are transcribed. Stress is not marked.

Syllabic and postvocalic r is transcribed [ɚ]. Syllable-initial r is transcribed [r]. Hence, the word encumber is transcribed [In-kəm-bɚ], but encumbering is transcribed [In-kə-bə-rIŋ].

Syllabic lateral and nasal resonants are transcribed with a schwa [ə] preceding them ([bæt]̩] battle > [bæt-əl]), so that each syllable is justified with a vowel.

Glottal stops are transcribed as the phonemic symbol with which they are associated. For example, the strictly phonetic [bəʔn̩] button is transcribed [bət-ən].

Flapped r's are also transcribed as idealized pronunciation would suggest. That is, the phonetic [læřɚ] is transcribed [læt-ɚ] latter, but [læd-ɚ] ladder.

Tense vowels, which tend to be diphthongal, are not transcribed as diphthongs, but rather as [i,e,u,o].

Frequently occurring suffixes such as -ing or ed, are transcribed with preceding consonants as part of the final syllable. This feature enables the user to retrieve more specific ing forms than a trivial listing all words ending in ing, although that, too, is possible, as explained below.

One major problem in entering the data of spoken English is representing the variant pronunciations of a given word. Although we have 62,500 words in the corpus written in English orthography, we have approximately 72,000

113

correspondences in phonetic transcription.   Examples of variant pronunciation entries include the following:

    stumbling: stəm-bə-lIŋ, stəm-blIŋ

    suggestion: sə-ǰĔs-čən, sə-ǰĔš-čən, səg -ǰĔš-čən,
                 səg-ǰEs-čən

    status: stæt-əs, stet-əs

    suspect: səs-pEkt, sə-spEkt

    syndicate: sIn-dɪ-kət, sIn-də-ket

    subject: səb-ǰEkt, səb-ǰIkt

    standby: stænd-bai, stæn-bai

    strictly: strIkt-li strIk-li

    stance: stæns, stænts

    caught: kat, kɔt

The data base has been entered on a Macintosh SE computer on which phonetic characters can be designed. The vocabulary base is a subset (54%) of the current WordPerfect Speller.

Developing a Solution: Computer Program

Robert Stevens' work is focussed on developing a program that will run as fast as possible with the data compressed to fit easily on two floppy discs.   It would be nice to display the phonetic notation on graphical display such as the MacIntosh computer provides.   The rhymer retrieval system is very versatile, enabling the user to retrieve data in a variety of forms, according to need, as the following examples illustrate:

Perfect Rhyme.   Poets speak of "perfect rhyme," which is essentially rhyming with the final vowel of a word, including diphthongs, plus any consonant(s) following it. This type of rhyme is elicited in the program by the function "End of word."

Identical Rhyme.   This type of rhyme is the same as perfect rhyme, except the consonant(s) preceding the vowel are also included.   This type of rhyme is elicited by the function "last syllable."   It retrieves all sounds following the last syllable boundary.   In some cases, the last syllable will also be a perfect rhyme.

Word Initial Rhymes.   The mirror images of perfect rhyme and identical rhyme are retrievable from the beginnings of

words. That is, the function "first syllable" will retrieve all sounds that occur before the first hyphen. The function "beginning of word" will retrieve the first vowel and any consonant(s) preceding it.

Alliteration. Alliteration is the successive occurrence of a particular consonant as the first sound of successive syllables. Alliterative consonants in specific positions in a word can be retrieved by selecting the desired consonant from the sound menu, and placing it in the desired sound pattern. For example, words with alliterative p's as in peppers or piper can be elicited by entering: p*-p*, where the asterisk represents any sound(s), and the hyphen represents a syllable boundary. The use of an asterisk in this manner is referred to as a "wild card."

Assonance. Assonance is the recurring of a particular vowel sound in successive syllables. Words with desired vowel sounds may be retrieved in the same manner as alliterative consonants were retrieved. For example, if the user were to retrieve words having the vowel [i] in both of the first and second syllables, he could enter: *i*-*i*, where the asterisk represents any sound(s) (or lack of sound), and the hyphen is a syllable boundary. This function would retrieve, for example, the word retrieve, and other words with the vowel [i] in both the first and second syllables.

Consonance. Consonance is the repetition at close intervals of syllable-final consonants. The same type of procedure as that used for alliteration and assonance may be used to retrieve repetitive consonants.

Other Sound combinations. It should be clear that with the use of the "wild card" feature, the user may retrieve any consecutive or nonconsecutive sound sequences. This feature makes the rhymer a very useful tool for linguists, teachers, and material developers.

Another appealing feature of the program is that the word on which the cursor is placed may be replaced directly with any of the rhyming words retrieved.

Demonstration of the Program

As part of the demonstration of the rhymer, it seems appropriate that we help "Dear Abby" with her problem of finding other words with the final "gry" sound. A quick check with the rhymer reveals at least the following:

    agree, degree, disagree, filigree, pedigree, and vinegary
    (There is also another pronunciation of vinegary that
    includes a schwa vowel between the [g] and the [r])

Some persons have asked for words rhyming with orange.

115

The rhymer reveals the following "perfect rhymes:"

hinge, syringe, astringe, singe, binge, cringe, fringe, tinge, impinge, infringe

(This paper was presented March 14, 1989 in the Deseret Language and Linguistics Symposium at Brigham Young University)