



PROJECT MUSE®

Reversing a One-Way Bilingual Dictionary

Leonard Newmark

Dictionaries: Journal of the Dictionary Society of North America, Number 20, 1999, pp. 37-48 (Article)

Published by Dictionary Society of North America

DOI: <https://doi.org/10.1353/dic.1999.0004>



➔ *For additional information about this article*

<https://muse.jhu.edu/article/458404/summary>



Reversing a One-Way Bilingual Dictionary*

Leonard Newmark

“One day we will go back to Kosov[a]. That’s our land.”
— Ramada Shaqiri, 30 March 1999

After completing a ten-year project to write an Albanian dictionary, published in March 1998 by Oxford University Press as the *Oxford Albanian-English Dictionary* and designed specifically for users who want to read Albanian and whose access language is English, I decided to prepare a companion English-Albanian dictionary for users who want to write Albanian, but I did not want to devote another ten years to that compilation. I wondered whether I could produce a useful bilingual dictionary with the reverse orientation by automatic conversion of the entries in the data files from which the first dictionary was generated. This paper is a report on the degree to which the attempt succeeded and the degree to which human intervention was required. Examples are provided to illustrate some rather surprising results, and a general conclusion is drawn for bilingual lexicography.

For languages of limited worldwide commercial importance, like Albanian, it seems particularly important to use computational techniques to derive new dictionaries from lexical data files compiled for some other purpose, especially if those files are extensive and have information otherwise difficult to come by in machine-readable form. The richness of the lexical data files from which my Albanian-English dictionary was generated is evidenced by that dictionary’s 75,000 entries and subentries, more than are found in any other dictionary of Albanian. Those files already provide a number of features that dis-

*This paper is a reworked and expanded version of the paper I presented at the 8th EURALEX International Congress (4-8 August 1998) and published in the proceedings of that congress.

tinguished this dictionary from many other bilingual dictionaries: 1) inclusion of large numbers of nonstandard items (marked by asterisks) as well as all attested standard stems; 2) marking of morpheme boundaries in Albanian words; 3) inclusion of some 16,700 phrasal expressions, in particular, phrasal names, collocations, idioms, and proverbs; 4) use of large numbers of bipartite definitions with a discursive description of the sense followed, after a colon, by English synonyms exemplifying that sense; 5) inclusion of large numbers of terms for grasses, flowers, birds, and fish with their scientific definitions; 6) inclusion of a modest amount of encyclopedic information to explain words whose strictly lexical meaning would not make their use in Albanian contexts intelligible; 7) listing of the various stem forms of lexemes as separate entries in their own alphabetical position to enable readers to decipher otherwise mystifying forms encountered in actual texts; 8) indication of the specific limits of variation that leave idiomatic senses intact (e.g., in phrasal expressions, marking a verb that can appear in any of its inflected forms by giving it in citation form with a symbol (•) at the end of the stem); 9) elaborate labeling of Albanian distinctions in domain and register; 10) rendition of phrasal expressions by stylistically similar English expressions, frequently supplemented by literal translations (enclosed in quotation marks) to enable more nuanced understanding.

Each entry in the plain text, flat data files from which the Albanian-English dictionary was generated is a line consisting of an Albanian word or phrase followed by a definition in English (or by a cross-reference to another line). Each line is embedded with simple visible two-letter formatting codes (e.g., HW [headword], DF [definition], TK [technical name], CO [collocation]) immediately preceded by a period (.) and immediately followed by what will get the formatting assigned by that code. The easily redefinable codes are later translated by a set of UNIX scripts into formatting instructions in TeX, which can go directly to a printer or indirectly by translation into Post-Script files.

The simplicity of such transparent and flexible coding for entering the data, in contrast with elaborate schemes requiring complex coding by experts into predefined structures,¹ was initially dictated by limits typical of languages that attract little commercial interest and

¹For example, those used in the architecture described by Willy Martin and Anne Tamm in "OMBI: An editor for constructing reversible lexical databases," *EURALEX '96 Proceedings* 1-2, 675-87.

thus little financial support. If a language researcher has been consistent in the use of other diacritical devices — line breaks, punctuation, capitalization, etc. — to enter lexical information into a data file, it is quite easy to transform that information automatically into files with the codes that I used, thereby gaining access to an entire formatting system to produce camera-ready copy for a publisher.

For example, for my Albanian dictionary the code “.HW” in the raw data file introduces a new line that will become a hanging indent for a headword in boldface; “.GI” indicates that the grammatical indicator that follows will be set in italics; “.DF” will automatically be assigned a number in proper sequence for the definition (in roman typeface) that follows when the file eventually appears in print; and “.ID” will be interpreted as a raised dot followed by a space and boldface up to the next period.

From a primary data file a Makefile (containing complex scripts for sorting, merging, and editing these files) on a computer using a UNIX operating system automatically alphabetizes the headword line:

```
.HW80lak.GInm.DFcurve.DFbight, loop.DFsnare, noose.DF-
twisted leather thong used for tying up moccasins.DF-
halter/saddle rope; stirrup loop.DFhorsehair bow for a bowed
lute.DFtendon at the rear hollow of the knee: hamstring.
DFknuckle.DFbowed declivity/pass in a range of hills/moun-
tains; area with many steep declivities
```

finds the phrasal expressions from another file:

```
.IDnuk i be6#n lak fjale6s.PDto keep one's word
.IDlak bore.PDsnowdrift, mound of snow
.IDlaku i qafe6s.UIContempt.PDrope wound around a woman's
neck and used to carry loads; symbol of a woman's subjugation
and hard work
```

and merges them and formats them into a TeX file as:

```
\SR{80}\HW{lak}\GU{nm}\DF curve\DF bight, loop\DF snare,
noose\DF twisted leather thong used for tying up moccasins\DF
halter\slash saddle rope; stirrup loop\DF horsehair bow for a
bowed lute\DF tendon at the rear hollow of the knee: ham-
string\DF knuckle\DFbowed declivity\slash pass in a range of
hills\slash mountains; area with many steep declivities
```

\ID{nuk i b\“e\VB\SC{n} lak fjaN\“es\}PU to keep one's word
 \ID{lak bore\}PU snowdrift, mound of snow
 \ID{laku i qaf\“es\}UI{Contempt\}PU rope wound around a woman's neck and used to carry loads; symbol of a woman's subjugation and hard work

These lines, in turn, are converted automatically by a TeX-based formatter into camera-ready printed pages (or into PostScript files for PostScript driven printers) with automatically numbered definitions and formatted double columns with headers showing the first and last words on the page (see Sample 1). Because this conversion is done only as the files are printed, last-minute additions, deletions, or movements of entries or their parts are automatically accommodated, requiring no renumbering or page adjustment by a human editor when the reversed dictionary is printed out (Sample 2).

For the projected English-Albanian dictionary, I used the formatting codes in the data files for the dictionary together with the commas (separating equal alternates) and semicolons (separating subsenses) in the definitional portion of the head entries, as crucial boundary markers for automatic search and replace operations (to exchange HW and DF markers) to create my first draft. In the beginning, I did not really know what to expect of such an automatic reversal process. Because the codes embedded for formatting purposes in that base automatically mark the boundaries of the parts of the entries, I knew that there would be no problem reversing single-word equivalents like **kolegjialite't** *nm* (*Bookish*) collegiality. Entries with strings of single-word equivalents for the Albanian word could be converted

Sample 1

PARTIAL PAGE OF THE ALBANIAN-ENGLISH DICTIONARY

lak

la'kishte

lak *nm* 1 curve 2 bight, loop 3 snare, noose 4 twisted leather thong used for tying up moccasins 5 halter/saddle rope; stirrup loop 6 horsehair bow for a bowed lute 7 tendon at the rear hollow of the knee: hamstring 8 knuckle 9 bowed declivity/pass in a range of hills/mountains; area with many steep declivities
 • **nuk i bē- lak fjalēs** to keep one's word
 • **lak bore** snowdrift, mound of snow
 • **laku i qafēs** (*Contempt*) rope wound around a woman's neck and used to carry loads; symbol of a woman's subjugation and hard work
lakialdre'dh- *vt* to wind [] this way and that
 • **lakadredh- fjalën** to take back one's words

lakër'o- *vt* to cut/tear [] into small pieces: shred, tatter, cut [] up; cut to shreds
lakër'o/he- *vpr* to become all tattered; get cut to shreds
laklēs *n* unreliable/shifty person
lakil'm *nm ger* 1 **lak'o-**, **lak'o/he-** 2 road curve 3 [*Ling*] grammatical declension
lakil'nē *nf* 1 saddle of a mountain/hill; precipitous mountain pass 2 winding road up a mountain
 • **lakmie'c** OR **lakmie's** *adj, n* (*Pej*) greedy (person); pig
lakmie'sē *nf* = **lakmi'**
lakmil' *nf* 1 (*Pej*) voracious desire; avarice, greed 2 inordinate fondness/passion 3 envy

Sample 2

SAMPLE PAGE OF THE PROJECTED ENGLISH-ALBANIAN DICTIONARY

quibble

- to get in line; queue **futet** *në rradhë*
- quibble**
 - to quibble, cavil **1 ngre• vogëlsira 2 ça• n! nda• n qimen (më) katërsh/dysh**
- quibbler**
 - person who quibbles over details *n* **buurrinla'r**
- quiche**
 - vegetable quiche made with corn meal *n/f* **nga'stër**
- quick**
 - fast, fleet; rapid; quick *adj (i)* **shpejtë**
 - in a hurry; hurried, quick *adj (i)* **nxltiu'alr**
 - quick as lightning *adj (i)* **vetatill'mitë, rrufe'shëm**
 -
 - quick; clever *adj (i)* **zhderivje'lliët**
 - very quick and clever, skittering quickly *adj* **shpuzia'k**
 - keen-witted in speech; articulate; eloquent; quick on the response *adj* **gojëbr'isk**
 - ready to do any work, quick to do work, unhesitant to work *adj (i)* **palpërtiu'eishëm**
 - mentally/physically quick; adroit, agile, deft; clever, smart *adj (i)* **shka'thiët**
 - very quick and agile constantly moving, restless *adj* **hardhiucia'n "like a lizard"**
 - skillful and quick in work *adj (Colloq)* **sikala'ç**
 -
 - quick and dirty patch, temporary fix *n/f (Colloq)* **mballo'më**
 - hasty repair/patch, makeshift patch; quick fix *n/f (Fig)* **pateri'cë**
 -
 - quick and clever *adj (Fig) (si)* **shkëndl'j!ë "like a spark"**
 - quick and adroit *adj (Fig) (si)* **ble'të "like a bee"**
 - very quick; short and quick; quick as lightning, like lightning *adj (si)* **rrufe "like lightning"**
 - quick as a flash, in an instant *adv (si)* **flakërlil'më**
- quicken** *vt* **1 shpejto•, përshpejto• 2 (Fig) gjallërio• n, ngjall•**
- quicklime**
 - quicklime, burnt lime **gëlqere e gjallë/pashuar**
- quickly**
 - quickly **me të shpejtë**
 -
 - in a hurry, quickly *adv* **fugil'mithi, ngalre'ndias**
 - in a sudden surge, quickly *adv* **gjyrr'kithi**
 - in a fast trot, quickly *adv* **ling, l'ngithi**
 - in quick movements (like a butterfly) quickly *adv* **1 fluturil'mithi 2 (Fig) (si) flu'tur**
 -
 - quickly, rapidly *adv* **shpejt**
 - speedily, quickly; in a hurry; immediately, right away *adv* **shpejtias OR shpejtiaz**
 - hurriedly, hastily, quickly *adv* **nxltill'mithi**
 - briskly, quickly *adv* **çapishpejtë**
 - fast as a bullet, quickly *adv (si)* **plumb**
 - in a flash, with a swish, quickly *adv (Colloq)* **fl'u**
 - on the double, in a hurry, quickly *adv (Colloq)* **tro'kithi**
 - right away, quickly *adv (Colloq)* **fët**
 -
 - in a very short time, very quickly, quick as a bunny **fët e fët**

quiet

- before the paint even dries: quickly, soon **pa u tharë boja**
- quickly, on the run **në këmbë e në dorë/duar**
- quickness**
 - speed; quickness; rapidity *n/f* **shpejtësi'**
 - agility, adroitness, nimbleness; smart, mental quickness, cleverness *n/f* **shkathitësi'**
 - agility, adroitness; quickness; cleverness; liveliness; expressiveness *n/f* **zhderivje'llitësi'**
- quicksand**
 - quicksand; deep mud *nm* **humbia'k**
 - quicksand *n/f (Fig)* **le're**
 -
 - quicksand **rëra të lëvizshme**
- quicksilver** *n/f (Colloq)* **engjërdh'i**
 - mercury, quicksilver *n/f (Chem)* **zh!vë (Hg)**
 -
 - a mind like quicksilver **mendje e prehtë**
- quick-tempered**
 - hot-blooded, quick-tempered, irascible *adj* **gjakinxe'htë**
 - quick-tempered, touchy *adj* **nevril'k**
- quick-witted**
 - quick-witted, bright, clever *adj* **mendije'shka'thiët**
 - quick-witted, sharp, smart, clever *I § adj* **mendijel'ho'llë II § adj (i) kujtlu'eishëm**
 - clever on the uptake, sharp as a tack, quick-witted *adj* **gojëmpre'hitë**
 - sharp as a tack, quick-witted *adj* **fjalëmpre'htë**
 -
 - not quick-witted, unintelligent *adj (i)* **paikujtlu'eishëm**
 -
 - clever, quick-witted child *nm* **ça'tër**
- quick-wittedness**
 - quick-wittedness, sharpness, cleverness *n/f* **mendijel'holësi'**
- quid (of chewing tobacco)** *nm* **shuk (duhani)**
- quiet**
 - quiet, silent *adj (i)* **he'shtlur**
 - quiet, calm, still *adj (i)* **qetë**
 - noiseless, quiet, silent *adj (i)* **paizhu'rmë**
 - not noisy; quiet *adj (i)* **palndi'eir**
 - staying out of trouble, quiet *adj (i) (Colloq)* **paishe'rr**
 -
 - well-behaved; gentle, calm, quiet *adj (i)* **u'rtë**
 - cool, temperate, calm, quiet, sober *adj* **gjakfto'htë**
 - muted, muffled, dull; quiet, silent *adj (i)* **shurdh'iër**
 - taciturn, quiet, silent *adj* **gojërru'dhiur**
 - comfortable, restful; quiet and relaxed, peaceful; cozy *adj (i) (Colloq)* **reha'tishëm**
 -
 - furtively, sneakily; secretly, on the quiet; imperceptibly *adv* **ti'nëz**
 - keeping quiet and well-behaved (under pressure by an authority) *adv (Fig)* **sus**
 -
 - quietness, quiet; calmness, calm; tranquillity *n/f* **qetësi'**
 - peace and quiet, relaxation, calm *n/f (Colloq)* **rehatil'**
 - repose, serenity, quiet, calm *n/f* **ta'shë**
 -
 - to keep silent, be quiet *vi* **hesht•**
 - to be silent/quiet, say nothing *vi* **shu'•n**
 - to be still/quiet, stop (talking) *vi* **më'n**

automatically by duplicating the entry for each equivalent, while moving that equivalent into head position. Thus, an entry — **skërly'ell** *nm* [*Anat*] *trachea*, *windpipe* — could automatically be duplicated, reversed, and then included as identical subentries, once under the head entry **trachea** and then again under the entry **windpipe**:

trachea

trachea, *windpipe* *nm* [*Anat*] **skërly'ell**

...

windpipe

trachea, *windpipe* *nm* [*Anat*] **skërly'ell**

I was surprised and happy to discover that such subentries incidentally (and accidentally) offer useful information that is not apparent in bilingual dictionaries compiled in more traditional ways: the Albanian equivalent serves neutrally both as a specialized word *trachea* and as the ordinary word *windpipe*, the fact of that neutrality indicated automatically by the equation of the single Albanian word to the two English equivalents.

It seemed unlikely, even absurd, however, to expect success for entries with more complex multiword definitions. And indeed, explanatory entries — **my'kë** *nf* *blunt side of a bladed tool* — defied automatic conversion, because none of the component words of the English explanation could be translated as *my'kë*. However, emboldened by the success of the conversion of the strings of single-word equivalents, I tried a modification. For these multiword entries, I first marked one or more of the words in the explanation as index words, and in a few cases I decided to attach an index word that was not part of the explanatory definition. From that point on, the reversal of the entry and placement under that index word was automatic, using the UNIX scripts developed for the Albanian-English dictionary to place phrasal entries under appropriate headwords and to give them consistent TeX formatting. Thus, having marked *blunt* as an index for the explanatory Albanian-English entry, I could now automatically generate under the head entry **blunt** the useful subentry:

blunt side of a bladed tool *nf* **my'kë**

At this point the new dictionary began to look like a dictionary thesaurus, not an unwelcome development for a language like Albanian for which no other thesaurus exists, nor is likely to exist otherwise. (I will return to this point later.)

The next set of at-first-sight intransigent entries consisted of the thousands of phrasal entries in the Albanian-English dictionary. Here again, however, the index-word technique produced the interesting and useful result of greatly enriching this bilingual dictionary: 1) in many cases, it provided examples and context for English and Albanian matching; and 2) it automatically added a fairly large number of phrasal entries that are missing from any other English-Albanian dictionary. Thus, the automatically generated entry **jinx** now appears with an automatically generated, rich set of subentries²:

²No entry for **jinx** appears at all in any existing English-Albanian dictionary. Here is the list of those dictionaries (in descending order of comprehensiveness).

Gasper Kiçi and Hysni Aliko, *English-Albanian Dictionary* (Second Edition, privately printed in Italy, 1991, 27,000 headwords, 627 pages) offers more phrasal definitions than any of the others.

Ilo Stefanllari, *Fjalor Anglisht-Shqip* (Tirana: Shtëpia Botuese 8 Nëntori, 1986, 442 pages) is widely used in Albania.

Stuart E. Mann, *An English-Albanian Dictionary* (Cambridge: Cambridge UP, 1957, 21,000 headwords, 434 pages) is especially rich in bird and plant names. Its central Gheg dialect forms of Albanian are considered nonstandard in present-day Albania, and its stoutly British focus makes many of its entries unintelligible to present-day American users.

Ministria e Arësimit dhe e Kulturës, *Fjalor Anglisht-Shqip* (Tirana: Mihal Duri, 1966; reprinted in Prishtina, Yugoslavia in 1969 and 1972 with the title *Fjalor Anglisht-Shqip për Shkollë e Mesme*, 20,000 headwords, 340 pages) is inferior to the Stefanllari dictionary.

C.A. Chekrezi, *English-Albanian Dictionary* (Boston: Ilia Chapullari, 1923, 13,000 headwords, 187 pages), in a tiny format, has enjoyed considerable popularity among older Albanian-American users.

Nelo Drizari, *Albanian-English and English-Albanian Dictionary* (New York: Frederick Ungar Publishing Co., 1957, 184 pages in the English-Albanian section) is full of inaccuracies.

Ylvi Basha and Muhamet Kapllani, *Fjalor Themelor Anglisht-Shqip* (Tirana: Shtëpia Botuese e Librit Shkollor, 1972, 2,500 headwords, 74 pages).

I. Stefanllari and V. Dheri, *Fjalor frazeologjik Anglisht-Shqip* (Tirana: Shtëpia Botuese 8 Nëntori, 1980, 389 pages) is a hodgepodge of old literary quotations ("to give somebody a Roland for an Oliver") and modern colloquial expressions ("to give somebody credit for"), without distinguishing the two and with Albanian glosses not usable as equivalents.

Sadik Berisha, *Fjalor Praktik Drejtshkrimor dhe Drejshqiptimor Anglisht-Shqip* (Prishtinë: Enti i Teksteve dhe i Mjeteve Mësimore i Krahinës Socialiste Autonome të Kosovës, 1984, 10,000 words, 239 pages) is just a list of English words without definitions, to be used as a speller.

jinx *n* (*Colloq*) *ters*

1. *jinx* who attracts bad luck by mentioning it *nm* *gojē|te'rs*
2. a terrible *jinx* (*Crude*) *Tersi i Pojanit*
3. to be a *jinx* ("to have one's leg wrong") *e ka• këmbën të prapë*
4. to be afflicted by the evil eye: be bewitched, have a *jinx* *vpr*
për|sy'ish•et
5. to rid [] of the evil eye: break the *jinx* on [], release [] from a curse
vt [] *ç|më|sy'ls•*
6. to *jinx* <>, bring bad luck to <> <>*vē'• çullin*

The head entry *jinx* includes a symbol *n* specially defined as implying that the Albanian *ters* is a noun stem that is masculine in this form, but having a corresponding feminine form *te'rse*. The entry includes the register label "(Colloq)" — no innovation in itself, of course; the important point is that the elaborate number of informative labels and their frequency of use in the Albanian-English dictionary are captured at no cost in this derived English-Albanian dictionary.

Subentry 1 illustrates the use of internal morpheme boundaries in Albanian words by the symbol |, together with the indication of primary stress by the symbol ' after a vowel letter: such indication helps the user to pronounce the Albanian word as well as to understand its structure — in this example showing that *gojē|te'rs* is a compound whose second element is *ters*.

Subentry 2 shows that vulgar expressions, marked as "Crude," are included in this dictionary because they are generated automatically and for free. If I were compiling the dictionary from scratch, I would not have sought out such an entry, because the corresponding English expression is of no particular value and would not have called for translation.

Subentry 3 illustrates the use of the symbol • at the end of a verb stem to indicate that the verb (*ka•* 'to have' in this example) may appear in any of its inflected forms, corresponding to the natural use of the infinitive form in English to indicate the variability. Thus, 'Nexhmija was a *jinx*' may be rendered in Albanian by *Nexhmija e kishte këmbën të prapë*, with the past form *kishte* maintaining the idiom in place of the stem *ka•*. This entry also illustrates the frequent use of supplementary literal translations in phrasal expressions, characteristic of my Albanian-English dictionary but not characteristic of many bilingual dictionaries compiled in traditional ways.

Subentry 4 illustrates a frequent form of definition in the Albanian-English dictionary: a bipartite definition with a discursive

description followed by one or more approximate English equivalents. The grammatical label “*vpr*” (passive/reflexive verb) tells the user how the verb is to be conjugated; the location of • in the form **për|sy|sh•et** shows where the verb stem ends, and the italics after • show that the *e* (marking *vpr* forms) and *t* (marking 3d person singular) are evanescent, in the sense that they do not appear throughout the conjugation. In the form **my'kĕ** cited above, we saw a similar use of this typographical convention to indicate that the *ĕ* is evanescent: for example, it does not appear in the nominative definite form **my'ka**.

Subentry 5 is again in the form of discursive description plus English equivalents. It illustrates how [] in the English expression stands for a pronoun in the accusative case, which in turn may stand for any accusative object. The placement of the corresponding [] in the Albanian construction tells the user where a corresponding accusative pronominal clitic³ will be if the *vt* (transitive verb) in Albanian is to sound natural.

Subentry 6 connects the lemma item **jinx**, using a comma, to a synonymous, somewhat less colloquial expression in English, indicating the tonal range of the corresponding Albanian expression. It also illustrates how <> in the English expression stands for a pronoun in the dative case, which in turn may stand for any dative object. The placement of the corresponding <> in the Albanian construction reminds the user where a corresponding dative pronominal clitic will be if the verb is to sound natural in Albanian.

These examples illustrate the results of reversing the dictionary by automatic procedures. The innovative characteristics of the original dictionary and consequent value to the user of the reverse dictionary of the information that flows from the automatic (actually semi-automatic, considering the intervention of the editor in choosing index words) reversal would likely be lost if the dictionary had to be compiled from scratch, in large part because that value would not repay the labor necessary to include them, particularly for languages like Albanian with limited commercial appeal. The richness of concatenational information in the reversed phrasal entries for **jinx** is a natural consequence of the way in which the entries were generated automatically from the source dictionary, in striking contrast with the conven-

³The rules governing the presence of pronominal clitics in Albanian are quite complex. When the clitics do appear, however, they precede the verb, except optionally when the verb is in the imperative mood.

tionally compiled English-Albanian dictionaries, whose editors worked from a purely Albanian perspective or from uncontextualized English word lists.

In addition to these valuable features of the source dictionary, the semiautomatically reversed dictionary offers — with little cost in additional labor — some 2,000 names of birds, flowers, grasses, fish, diseases, etc., followed by their Latinate scientific names.

shark

shark [*Ichth*] **qen deti** ('sea dog')

angel shark *nf* [*Ichth*] **skadhi'në** *Squatina squatina*

angular rough shark *nm* [*Ichth*] **peshklalqen-de'rr** *Oxynotus centrina*

bramble shark *nm* [*Ichth*] **yll'za'k** *Echinorhinus brucus*

gray shark *nm* [*Ichth*] **peshklalqe'n** *Galeorhinus galeus*

great blue shark *nm* [*Ichth*] **peshklalqe'n** *Prionace glauca*

A major value of the derived dictionary lies in the way in which so many of its entries and subentries automatically provide contexts (in English) for choosing among alternative Albanian translations of an English term.

windpipe

windpipe, larynx *nm* (*Colloq*) **gurma'z**

windpipe, throat *nm* (*np* ~*j*) (*Colloq*) **gu'rgull**

windpipe, trachea *nm* (*in phrasal expressions*) **rryl**

trachea, windpipe *nm* [*Anat*] **gabzhe'rr**

trachea, windpipe *nm* [*Anat*] **gërlla'c**

trachea, windpipe *nm* [*Anat*] **skërly'e'll**

larynx, windpipe *nm* [*Anat*] **lari'ng**

Contrast the value to a writer of an entry like this with the value of the entry for *windpipe* given in the two best English-Albanian dictionaries available at present:

windpipe (uind pajp) *n.*, skërlyell, rryl, gabzherr⁴

windpipe (windpaip) *n.* kanali i frymëmarrjes, gabzherr, laring⁵

⁴Kiçi and Aliko, see n. 2.

⁵Stefanllari, see n. 2.

Of course, these two comparison definitions were written for Albanian users rather than for English-access users; but the point of the comparison remains: the amount of information contained in the automatically generated definitions far exceeds the amount contained in either or both of the hand-generated definitions. Even Albanian users, as well as the English-access users for whom the dictionary was originally aimed, will find value in the sheer wealth of information it offers.

While the automatic and semiautomatic compilation of entries does save an enormous amount of human energy and has some valuable consequences, there remain many editing tasks for the human editor. Human judgment is still needed to decide whether particular apparent redundancies should be expunged or left in; for example, should the three mechanically generated subentries

trachea, windpipe nm [Anat] **gabzhe'r**
trachea, windpipe nm [Anat] **gërlla'c**
trachea, windpipe nm [Anat] **skërlfy'e**ll

be allowed to stand, should they be coalesced, or should one or more be eliminated as merely confusing to the user? For items that happen not to have been generated at all by the computerized methods, one needs to find those items and to decide which are important enough to be added. For example, no entry for *abattoir* nor for *abbreviate* was generated from the source dictionary. I discover their absence by comparing my word list with those of other dictionaries; choosing which dictionaries to consult also requires human judgment. After judging that the first word is dispensable, but the second is not, I still need to compose the new definition by hand, with the help of an Albanian colleague.

This part of the process will parallel that used for the Albanian-English dictionary: I consult all the word lists and definitions in other English-Albanian dictionaries and construct tentative definitions that account for the variation in those dictionaries; then, for each definition, my Albanian colleague and I compare likely scenarios for the object word or phrase, to find matches and mismatches in the use of the term that require modification or replacement of the tentative definition. This process was surprisingly effective for the Albanian-English dictionary, as evidenced by subsequent (and continuing) checks of definitions against token occurrences in published running text, so I am optimistic about its effectiveness for the new dictionary. Preparatory electronic corpus-based lexicographic study is premature for Al-

banian, in view of the paucity of available digitized material in the language. The problem is even greater for most other noncommercial languages, and that makes the methods used in the current project worthy of consideration for those languages.

As may be apparent from even the few examples in this article, the semiautomatically generated reverse dictionary ignores the boundaries between ordinary dictionary, thesaurus, dictionary of collocations, lexicon of idioms, technical dictionary, and encyclopedia. Judging from the result, this overstepping of bounds constitutes a strength rather than a weakness. From the point of view of the user, the dictionary gains in value by including a rich mix of information related to its words.

I have attempted to show a way in which a dictionary designed for one purpose can be exploited in an economical and unusually valuable way to produce another one that serves a very different purpose. The lexicographer need not worry about the formatting of the final product; it's enough to be consistent in demarcating the parts of the information entered. What form the actual dictionary takes can be worked out later; my elaborate final formatting is flexible and available if someone wants to use it for different data. The formatting actually turned out to be unexpectedly useful in producing a reverse dictionary with considerable thesaurus attributes.