Gen Dobry! 30 August 2000

****

INDEX

****

*** WELCOME! - ***
to the second issue of GEN DOBRY!, the e-zine of PolishRoots(tm). If you missed the
first issue, you can read it at:

http://www.polishroots.org/gendobry/GenDobry_vol1_no1.htm

We had unforeseen problems with distributing the first issue. We prepared it in pure text
format, but for some reason it went out as an attached Microsoft Word file, which some
recipients had trouble opening. Gary Mokotoff, publisher of the fine Jewish genealogical
publication Avotaynu

[http://www.avotaynu.com] and their e-zine "Nu? What's New?", told us about a better distribution system we might use. Webmaster Don Szumowski took a look and decided it should make this easier for you to download and read. It should also present Website addresses in a form you can click on -- much more convenient than what we did in the first issue. We hope you'll have no trouble with this issue, and we thank Gary for his generous advice.

Also, we started out planning to do Gen Dobry! on a quarterly basis. But Gary suggested that a monthly version would not be much more work than a quarterly. Besides, what sense does it make to note down timely info if you're going to wait three months to tell people about it? The more I thought about it, the more I realized he was right.

So we're going to try to do this on a monthly basis. There are no guarantees I can keep it up -- I have a lot of other commitments. But I'm willing to try.

You can help. Send me addresses of sites you've found that helped you. Send me ideas for articles, or, better yet, articles! Send comments, suggestions, anything you think of that might help someone else. Frankly, the less I have to write, the more I can focus on editing Gen Dobry!, and the better it will be. Other voices need to be heard, yours among them. If you have something to contribute, or just something to say, please E-mail me at <WFHoffman@prodigy.net>.

****

*** POLISH TRANSLATION SOFTWARE - WISHFUL THINKING? ***

by William F. Hoffman <WFHoffman@prodigy.net>

As one who has done a fair amount of translating -- though unfortunately I don't have time for it any more -- I've always been interested in the prospects of software that would do good translations, especially Polish to English and English to Polish. Furthermore, a number of folks have contacted me, asking if there was software that could help them with their translation needs. So I know there's interest in this question, and I thought it might be worth discussing in Gen Dobry!

Let's answer the obvious question first:

IS THERE SOFTWARE THAT DOES A GOOD JOB TRANSLATING POLISH TO ENGLISH AND VICE VERSA -- AS GOOD A JOB AS A COMPETENT HUMAN TRANSLATOR?

I'm afraid the short answer is, No. At least I haven't found any. I have doubts we'll see such software any time soon, for reasons I will discuss in due course.

Now, I willingly admit I haven't tried every program available. I couldn't afford to buy them all, even if I wanted to. So there may be a program out there that does a darned

good job. If you know of one, or if you use a program I discuss and feel I've underrated it, PLEASE write and tell me
about it.

I have looked at several packages, but as I searched the Web I found some I hadn't heard of. The Website addresses I found are given below, and I offer my best insights as to the apparent value of the software. You are welcome to try any of them that seem promising. But please remember: "Caveat emptor" ("Let the buyer beware"). I DO NOT, AND CAN NOT, MAKE ANY
PROMISE THAT THIS OR THAT PROGRAM WILL BE WHAT YOU WANT.

Before looking at specifics, however, let's discuss something very basic.

- WHAT DEFINES GOOD TRANSLATION SOFTWARE? -

My experience suggests this question must be answered on an individual basis, because a product that satisfies you might disappoint someone else. We can make general statements, however, that are reasonably accurate. I would say for any user the criteria for good translation software answer depend on three factors: 1) price, 2) your expectations, and 3) what language are you talking about?

First, price. This one's obvious. Clearly we all want the best product we can get at the lowest price. All things being equal, we'd expect the more expensive programs to do the better job. That doesn't necessarily hold true every time, of course. But it takes money to develop and test
software, so higher-priced products generally perform better than cheaper ones.

Second, expectations. Do you want a product that will take your English and turn it into natural-sounding, smooth Polish (or vice versa), without any effort on your part, other than pressing the right button on your mouse? If so, I feel certain there's no such critter, not yet. But if
you'd be happy with a program that takes normal English/Polish text and translates it well enough that you get the gist of what it says -- even if it's a bit garbled in places -- there are affordable products already available that can attempt that.

Finally, language. All languages are not created equal; some tend to express things in ways we comprehend more readily than others. Most people who grow up speaking English find Spanish, for instance, easier to learn than Japanese. Many Americans find German, for instance, enough like English that they can make their way through German records.

I'm afraid most English-speakers have a tough time with Polish, however. Not only do you have to find the right word, you have to put the right endings on nouns, adjectives, pronouns, and verbs. If you can't do that, you run a real risk of baffling any Pole you write to. Polish sentences
are knit together by the endings on words. English just doesn't do things that way; it relies

a lot on word order. Scramble the word order and we can't tell if man bit dog or dog bit man. In Polish, no problem -- the endings tell you who did what to whom, so word order is more flexible.
Good news for Poles, bad news for us.

Also, the hard truth is, there's more of a market for some languages than others. German, French, and Spanish get a lot of attention from developers of translation software because there's more of a market for them (or at least that's the perception). Polish, while not ignored, is farther down on the list. The key is, it takes money to develop software, and who's going to provide the up-front financing for really good Polish/English translation software? Frankly, I expect the Poles will do it long before anyone in the English-speaking world gets around to it, because they have more financial incentive.

So if you want software that will translate Polish perfectly for less than $100, good luck! If you want a program in that price range that will give you the gist of a Polish text, your chances are somewhat better. It depends to a great extent on how much you expect and how much you're
willing to spend.

- SOFTWARE AVAILABLE NOW -

There are several packages out there which promise you help with Polish for less than $100. Here are some Websites that offer these programs:

http://www.zem.co.uk/mlsoft/index.htm
http://translationsoftware.aimhi.com/software.html
http://www.bgtv.com/translator/software/polish.html
http://www.gy.com/language/po_trans_dic.htm
http://www.multilingualbooks.com/trans.html

They offer a variety of affordable products, and I'm sure all of them have their uses. I suspect, however, that they fall short of what the average user would want from a translation program. I say this because close reading of their descriptions suggests these programs take the "electronic dictionary" approach.

Such programs typically take a passage one word at a time, find that word in a dictionary database, and substitute the Polish word[s] given as its equivalent. The result is a literal translation, which may include several possible equivalents for terms with more than one meaning. In theory this is a reasonable approach. In practice, you may find that these programs
help you, but they don't really translate, as I will discuss in a moment.

I also note that at least one of the products, "Universal Translator 2000," has OCR [Optical Character Recognition] capability. That means you can scan a document (if you have a scanner!), turn it into text, and translate it. This is plausible; OCR software is available that does a very creditable job at a reasonable price -- but only if the text being

scanned is in clear print form. Please don't delude yourselves that you'll be able to scan, OCR, and translate photocopies of the scribbled documents normally involved in genealogical research! Leaving translation issues aside for a moment, OCR'ing handwritten text in English from an original is iffy even for high-priced programs; photocopied scribblings in Polish, German, Latin, or Russian -- forget it!

Another site worth a look is:

http://www.tranexp.com/win/NeuroTra.htm

I'm a bit suspicious in that this site is full of claims, but short on details, including price. The approach they say the program takes could produce good results, if implemented properly. It goes beyond the "electronic dictionary" approach, and that's promising, if true.

- OK, WHAT'S WRONG WITH ELECTRONIC DICTIONARIES? -

Nothing, if that's what suits your needs. I have the Collins YDP Multimedia Polsko-Angielski Angielsko-Polski Sl~ownik [http://www.ydp.com.pl] installed on my computer, and use it almost every day. It helps me think of that word that's on the tip of my tongue, or lets me check to be sure the word I've found is the one I meant. It helps me a lot. But I use it as a handy substitute for a bound, printed dictionary. I do the translating, it just helps me do it better.

The problem with the whole "electronic dictionary" approach is the same one you get when you consult a hard-copy dictionary. For many words there are lots of choices -- the bigger the dictionary, the more the choices -- but no way to tell which is the right one. Thus, for example, looking up "check" in a typical electronic dictionary will give you, among others, the translations _czek_, _rachunek_, _szach_, _sprawdzac~_, _powstrzymywac~_, etc. These words (and others, too!) all mean "check" in some sense.

The crucial info you need to know is that _czek_ is what you write to cover a purchase, _rachunek_ is your check or bill at a restaurant, _szach_ means "Check!" when you're playing chess, _sprawdzac~_ means "to check" in the sense of "to compare lists or other items for accuracy," and _powstrzymywac~_ means "to check" as in "We need to check inflation."

This goes to the heart of the matter. To use a dictionary well, you already need to have at least some knowledge of the language, enough to recognize and choose just the right word. We need translation software precisely to help us choose from among the possibilities -- to keep us
from annoying waiters in restaurants by warning them their King is in check, when what we mean to do is ask for the bill. Giving us more choices only aggravates the problem. It's as if you say, "I can't eat liver, I have trouble digesting it," and someone replies helpfully, "Oh, then let me give you a larger portion!"

Good translation software would need a stout dictionary database to consult, no question. But by itself that's not enough. The program still has to cope with the many different forms dictated by Polish grammar. What's more, it has to be able to deal with context, which is all-important. After all, "bitch" is a perfectly good English word... if you're talking about a female dog!

- THE ELECTRONIC DICTIONARY APPROACH: SAMPLES -

I noticed online some recommend using "InterTran" as a translation aid. InterTran is a no-frills example of the electronic dictionary approach. It's free, easy to access and use. It lets you key in a passage you want translated, or a Website address, and then translates the text you've
keyed in or the text that appears on the Website you specified.

I visited the InterTran site:

http://www.tranexp.com:2000/InterTran?

I decided to try something simple and see what it produced. I keyed in a well-known phrase from a song from "My Fair Lady." InterTran promptly told me this is how you say it in Polish:

Ten deszcz w Hiszpania zostaje gl~o~wnie w ten prosty.

Just for fun, I keyed in that Polish sentence and asked for an English translation. I got this:

This rainfall on Spain stays chiefly on this straight.

So much for "The rain in Spain stays mainly in the plain." It has turned into a comment on a wet Spanish heterosexual!

To be fair, the first part of the translation would at least be comprehensible to a Pole, although he'd wonder why you inserted _ten_ (Poles don't use the definite or indefinite article) and he might wince at "w Hiszpania" instead of "w Hiszpanii." But the program couldn't distinguish "plain," a broad, level expanse, from "plain," unadorned, open, obvious, clear. The Polish adjective _prosty_ is a good rendering of the latter. What I needed in this context was a form of the noun _równina_. The program chose the wrong word, and that rendered the whole sentence incomprehensible.

But after all, this sample may be too short, and thus not a fair test. Any good translator knows it's easier to do longer pieces because they give you more contextual clues, which suggest the right renderings. I decided to find a Web page in Polish and test InterTran with that. I found an
interesting page on Irish first names -- for those of you with Irish and Polish blood, it's at <http://www.kern.px.pl/imiona.html>.

I keyed this URL into the into the InterTran interface. I concentrated on two sentences. If you visit the site, they are the ones beginning "W praktyce formalne nazwiska..." and the one right after that,"Od kilku wieków..." I would translate them as follows:

"In practice formal names were used only on official occasions. In everyday use people were referred to either with the help of their parents' names -- e.g., Tadhg Johnny, Tadhg son of Johnny; Jimin Mháire Thaidhg, Jimin son of Mháire (the daughter of) Tadhg -- or with the aid of
bynames (or nicknames), e.g., Paidí an Ghleanna, Paidí of the valley, and Una Bhá, Blond Una.
"For many centuries English surnames have been generally used in Irish linguistic environments, and at times they are even more common than their 'official' Irish equivalents."

Here's what InterTran came up with:

"In practice shape surnames uses [sa~] only with z stiff pick-up. To what dies bids yourselves of [ludziach] or using [imion] bears, [np]. [Tadhg] [Johnny] [Tadhg] son [Johnniego], [Jimin] [Mháire] [Thaidhg] [Jimin] son [Máire] ([be~da~cej] [córka~]) [Tadhga] or using [przydomków] (or [przezwisk]), [np]. [Paidí] [an] [Ghleanna] [Paidí] ex valleys, [Una] [Bhá] [Bial~a] (fair-haired) [Una].
"Ex several century [powszechne] [sa~] on [s~rodowiskach] [irlandoje~zycznych] surnames english; downtime [sa~] they yet [cze~stsze] than them "[oficjalne]" correspondent irish."

You see the problem. Many Polish words were not translated correctly; I can't imagine how it got "bears" from _rodziców_, "of the parents." Many words were simply repeated in brackets -- these are terms the program couldn't recognize. I don't fault it for missing hard ones like
przydomków, "by-names." But how could it not catch _imion_ as genitive plural of the word for "first name"? It didn't even recognize _córka~_, a form of _córka_, "daughter." And what kind of translating program doesn't recognize _sa~_, "are"? Perhaps President Clinton doesn't know what is is, but a translating program should at least recognize forms of "to be"!

I admit I'm being a little too hard on InterTran. It will help you get the gist of many texts, especially if you recognize the limits of the program and make due allowance. And it is free. Still, if this is the kind of results an "electronic dictionary" program can produce, count me out. And I must tell you, it is fairly representative of the programs I have tested.

- DO WE HAVE A WINNER? -

So what would a truly effective translation program have to feature? I'd say it needs 1) a large dictionary, 2) algorithms that recognize and deal with grammar (including all those case endings on Polish nouns and adjectives, and all those verb forms), and 3) some

ability to deal with
semantics in different contexts.

I have some hopes for "Intelligent English-Polish-English Language Translation System for Windows 95/98 or NT" at this site:

http://www.languagetranslator.com/software/intran/

The promo, at least, says all the right things. It claims to contain over 200,000 words with over a million different English/Polish grammar forms. It says it "utilizes numerous semantic algorithms and grammar rules to offer the highest possible quality of the language translation. It can
identify the meanings and grammar forms of words in the sequence to translate the whole sentence more accurately..." And it allows you to tell the system to translate text within the following contexts: "General text; Ekonomical [sic]; Legal; Chemical; Biological; Physical; Medical; Technical; Computer."

Now we're getting somewhere -- maybe. This is the kind of approach a truly useful translation program would have to take. And it is priced at $160.00 plus shipping and handling; not bad!

I was glad to see on the site "an unaltered text resulting from fully automatic language translation." I looked at the sample, and was fairly impressed. The translation is far from perfect -- no one would make the mistake of thinking a native Pole wrote it -- but a Pole would certainly
understand it. It's a respectable effort. I also noted these comments (I have not corrected the grammar):

"You do not have to have a slightest knowledge of the destination language. The program will do everything for you. Automatically with possibility to change word meanings
After the automatic translation is completed you can click on any word and change the meaning of a given word
for something else from the pop-up box. Manually: This mode is the slowest, but results in the best quality of the translation. "For the official applications, manual corrections are strongly
recommended."

I like the last couple of sentences because they're realistic. They're saying the program will do a respectable job, but for best results you should check it for errors. That's sound advice for any translator, any time. But... well, it's the same old problem: how do you check it for errors if you don't speak Polish? And come to think of it, why would you need to if "the program will do everything for you"?

So I have doubts even about this program. Still, the material on the Website is interesting. I wish I could afford to buy a copy and test it. Since I can't afford to do that (especially for software I don't really need myself), I can't recommend it. What I can say is, if their

promo can be believed, the developers asked the right questions when they designed it. How successfully they answered them remains to be seen. Caveat emptor!

THE FUTURE?

Will we someday see really top-notch translation software? 30 years ago I would have scoffed at the notion. But I also thought we'd never see a chess program beat the best grandmasters; yet that has happened. In both cases, what's needed is a way either to make a computer mimic the creative functions of the human brain, or to enable it to fake those functions convincingly and produce the same results. If a chess computer can beat Garry Kasparov, I, for one, cannot regard good translation software as a pipe dream.

I don't think it will happen soon, however, and especially for the less "commercial" languages such as Polish, simply because of the perception there's not enough money in it. It will take a lot of work, and money, to produce a program that can do as good a job as a competent human. This project is not like the one that developed the best chess-playing computers. A lot of companies and brilliant programmers invested tons of money and effort in that project, in hopes it would yield valuable insights into Artificial Intelligence. In other words, a championship chess program was a means to an end. It offered no financial rewards, but all hoped it would produce insights valuable in creating software that would have commercial potential. I've seen no evidence that the same kind of effort has gone into translation software.

As I said earlier, if really top-notch software is produced, it will probably be aimed at languages like Spanish and German, and perhaps Russian. Kind of short-sighted, in my opinion -- I think Polish is a better bet, given the relative health of the Russian and Polish economies. I believe Polish programmers -- and there are some fine ones! -- will make the breakthrough that helps us with this problem.

Also, don't forget, genealogical research involves dealing with a lot of archaic or non-standard terms seldom seen in modern Polish. Good Polish-English software will surely focus on business and legal contexts, not genealogical. Good luck finding any program that translates _komornik_ or _chal~upnik_ correctly!

So I'm not holding my breath. But I might be in for a surprise. This is a case where I'd love to be wrong!

- A FEW AIDS YOU MIGHT HAVE OVERLOOKED -

Before ending, I should mention a couple of things I've run across that might offer some practical assistance. There is an interesting Website that concentrates on helping with genealogical records, located at this address:

http://users.vvi.net/jack/poltrans.htm

It includes pages offering help with a number of different formats of vital records.

Also, I recently bought Transparent Language's "Learn Polish Now," Version 8, because it was $49 at Best Buy with a $20 rebate. I figured at that price I couldn't go wrong. In fact, the "Learn Polish Now" program is a pleasant surprise. It gives you the chance to learn a lot of common
expressions and hear them pronounced properly. It features dialogs and multimedia presentations on Polish cities written in good Polish, not the stilted "Where is my pencil?" crap so many beginning language aids bore you with. I was especially impressed by a window that identifies each word grammatically, helping you recognize the grammar behind the forms you see. Included was an inexpensive microphone you can attach to your computer, so you can imitate the pronunciation of words and compare your version with how native pronunciation.

It won't necessarily teach you to speak like a native. But if you use it well and consistently, I think it can help you develop basic conversational skills. Travelers who intend to visit Poland, and want to avoid being clueless when Poles speak, might find it worth a try. The price is quite reasonable.

To sum up, there are translation programs out there, and some of them are worth a look. If you consider buying one, be sure you take into account its price and your expectations. I hope this article will help you make an informed choice, and I welcome reader feedback!

****

*** LETTERS TO THE EDITOR ***

We received a number of notes from readers praising the first issue of Gen Dobry! It was very gratifying to read them, and we thank all who wrote.

One in particular provided information well worth excerpting. It's from Matthew Bielawa <BielawaM@sacredheart.edu>, one of the data coordinators of the Polish Genealogical Society of Connecticut/the Northeast [http://members.aol.com/pgsne2/]:

"I know that you wrote that there must be just 'jillions' of good resources on the web, but I still would like to take you up on your offer and add one to your superb list.

http://www.pi.pl - 'Polskie Informatory'

"From this home page, you can access all the województwa, and then at each województwo you can access each powiat, and from there each gmina. At each of these levels you can find links or information on many various topics, such as culture, schools, political structure, tourism, etc. You won't always find information for each topic for each location (I believe it's a newer site), but it's a good place to poke around in."

****

*** POLISH FALCONS CAPITAL FUND PROJECT ***

PolishRoots(tm) is attempting to raise money to preserve membership records of the Polish Falcons. The Falcons are a national fraternal organization dedicated to preserving our rich heritage while documenting and publishing the many achievements and activities of the Polish-American community and Poland. Chances are very good any Polish-American family had relatives who were Falcons.

PolishRoots(tm) has an opportunity to preserve the Falcons' membership records for researchers, but it won't come cheap! Its funding obligation is $10,000, to be earmarked for this and other future projects. If successful, PolishRoots(tm) hopes to make these records available for your use by Fall 2001.

For more information, visit this Website:

http://www.polishroots.org/falcons.htm

****

*** WORLD WAR I DRAFT REGISTRATIONS ***

Jan Zaleski, Editor of the _Polish Eaglet_, the Journal of the Polish Genealogical Society of Michigan [http://www.pgsm.org/], wrote to contribute information on World War I Civilian Draft Registration. Here is an excerpt from his note:

WWI Civilian Draft Registrations

Free Database

** can be searched by entering:
1) first and last name
2) last name only
3) birth city
4) birth state (e.g. "POLAND")

NOTE: This computerized database DOES NOT CONTAIN ALL of the WWI draft registrations. It ONLY CONTAINS ROUGHLY 5% (or about 1.2 million) of all the men who registered, including registrations from approximately 13% of all counties nationwide.

Access WWI draft registration database by clicking on this address:

http://www.ancestry.com/ww1drafts

****

*** LDS FHL CATALOGUE CD ***

Debbie Greenlee <daveg@mail.airmail.net> posted to POLAND-ROOTS-
L@rootsweb.com a note that recommended buying a copy of the LDS Family History
Library Catalogue CD ($5.00). It has the most up-do-date data available. You can get it at
this address:

http://www.familysearch.org/

Click on "What's New" and scroll down to: "The Family History Library Catalog can
now be purchased online" etc.

****

*** MORE USEFUL WEB ADDRESSES ***

www.members.home.net/daveobee~.htm
This site, the home page of Dave Obee, owner of Genealogy Unlimited, has information
in these topics: German Colonies in Volhynia, Germans from Russia Research
Marketplace, Another source for Ostpreussen researchers, Canada.

http://www.PolishExpress.com

Norman Pieniazek's home page offers help with Polish genealogical research and
translations.

http://www.lithuaniangenealogy.org
This site has a collection of information on Lithuanian genealogical research, including a
chance to join Lithuanian Global Genealogical Resources, which publishes a semiannual
journal, _Proteviai_ [Forefathers].

http://www.polartcenter.com
This is the Website of the Polart Center in Hamtramck, Michigan, a good source for all
kinds of Polish books, artwork, crafts, knick-knacks, etc.

http://express.irk.ru:8101/1000/fam/
This "Onomastikon: Russkiye Familii" has notes in Russian on various Russian
surnames.

http://russian.about.com/homework/russian/msub20.htm?once=true&
Offers a whole series of pages with information on Russian names.

http://www.fortunecity.com/victorian/wooton/34/austria/contents.htm
A. John Birkholz posted a note to the list POSEN-L@rootsweb.com information on this
site, "A Short History of Austria-Hungary and Poland," London: The Encyclopedia

Britannica Company, Ltd., 1914. He explains that the Polish history begins with chapter 23. He adds that the chapters are quite large, so they will take time to download.

http://maxpages.com/poland
Tina Ellis, Listowner of the Polish Borders Surnames list, recently posted a list of the new pages added to this Website since March. It is too long to repeat here, but if you haven't visited this Website, you really should! I can't imagine any researcher who wouldn't find at least a few
features of potential benefit.

****

To be added to or removed from this mailing list, go to this address:

http://www.PolishRoots.org/GenDobry_signup.htm

YOU MAY REPRINT articles from GEN DOBRY!, PROVIDED: (1) the reprint is used for non-commercial, educational purposes; and (2) the following notice appears at the end of the article: Written by [author's name, e-mail address, and URL, if given]. Previously published by GEN DOBRY!, Vol. 1, No. 2, 30 August 2000. PolishRoots(tm): http://www.polishroots.org/.

****

****