ACADEMIA | Focus on Linguistics

BEWARE THE *BABADZIAFER!*

W e talk about technology, lexicography, and long-forgotten word senses in Polish with **Prof. Włodzimierz Gruszczyński** from the PAS Institute of the Polish Language, where work is underway on the Electronic Dictionary of the Polish Language of the 17th and 18th Centuries.

ACADEMIA: **How did you write your doctoral thesis?**
WŁODZIMIERZ GRUSZCZYŃSKI: On a ZX Spectrum computer, something some people are still unable to fathom. The perseverance I had to show truly verged on lunacy. I wrote it in sections of 10 pages each, because that technological wonder had just 64 kilobytes of memory.

And then you had to save to a diskette?
A diskette, in the mid-1980s? No, to an audio cassette. And then there was a problem of whether I would actually manage to print the thing out anywhere. Fortunately my friends from the Institute of Computer Science at the University of Warsaw adapted the gigantic line printers used for the supercomputers to work with the ZX Spectrum.

In other words, what you had was essentially a typewriter?
Yes. A bit more intelligent, because text could be moved around, etc. It brings a smile to one's face when one recalls under what conditions one started out. Today systems are being built that recognize speech and transcribe it in real time. Our colleagues at Samsung are now working on simultaneous translation of telephone conversations. The science-fiction writer Stanisław Lem would not have come up with such things.

Is that more fascinating or horrifying?
I still find it fascinating, although I can see the dangers. But that is the way things are with every invention. The ability to split the atomic nucleus brought great benefits, but some have used it for bad purposes. The Internet means surveillance and God knows what else,

OLD POLISH LANGUAGE GOES ONLINE

but also a whole ton of amazing things that simplify our lives. For instance, digital libraries. I now cannot imagine a researcher in the humanities or any lexicographer not having access to such libraries from any location on Earth. I am riding a train, writing a dictionary, and I can access the holdings of the Kórnik Library. Amazing.

But young people spend so much time online that they are ceasing to live in the real world.
A new generation is emerging, and at the same time we are witnessing a technological-cultural revolution comparable to the one initiated by Gutenberg's invention, or even surpassing it. Back then there were also people raising the alarm: "Our knowledge has always been inside us, but now it is beside us!" And indeed, the late fifteenth century people stopped needing their own memory so much, but increasingly made use of "external disks" instead – in other words, books. I re-

cently talked to an educated and intelligent businessman from Ghana and I asked him in what language literature is written in his country. "For us, the oral literature is more important. Everyone tells stories in their own ethnic tongue," he said. Which is much like back in Homer's times, and that is beautiful. But it is not their tradition that will prevail in the confrontation against the European and American tradition. So we feel horrified nowadays to see young people with their noses glued to their smartphones? The same worry was felt by parents and grandparents back in the 15th-16th centuries, seeing youngsters with their noses stuck in books rather than actually listening to their elders.

But then one at least had to know where to look for knowledge. There were no search engines.
Yes, but let's not delude ourselves that everyone read thick tomes cover-to-cover. For example, Queen Sophie's Bible consisted of fragments, selected for one
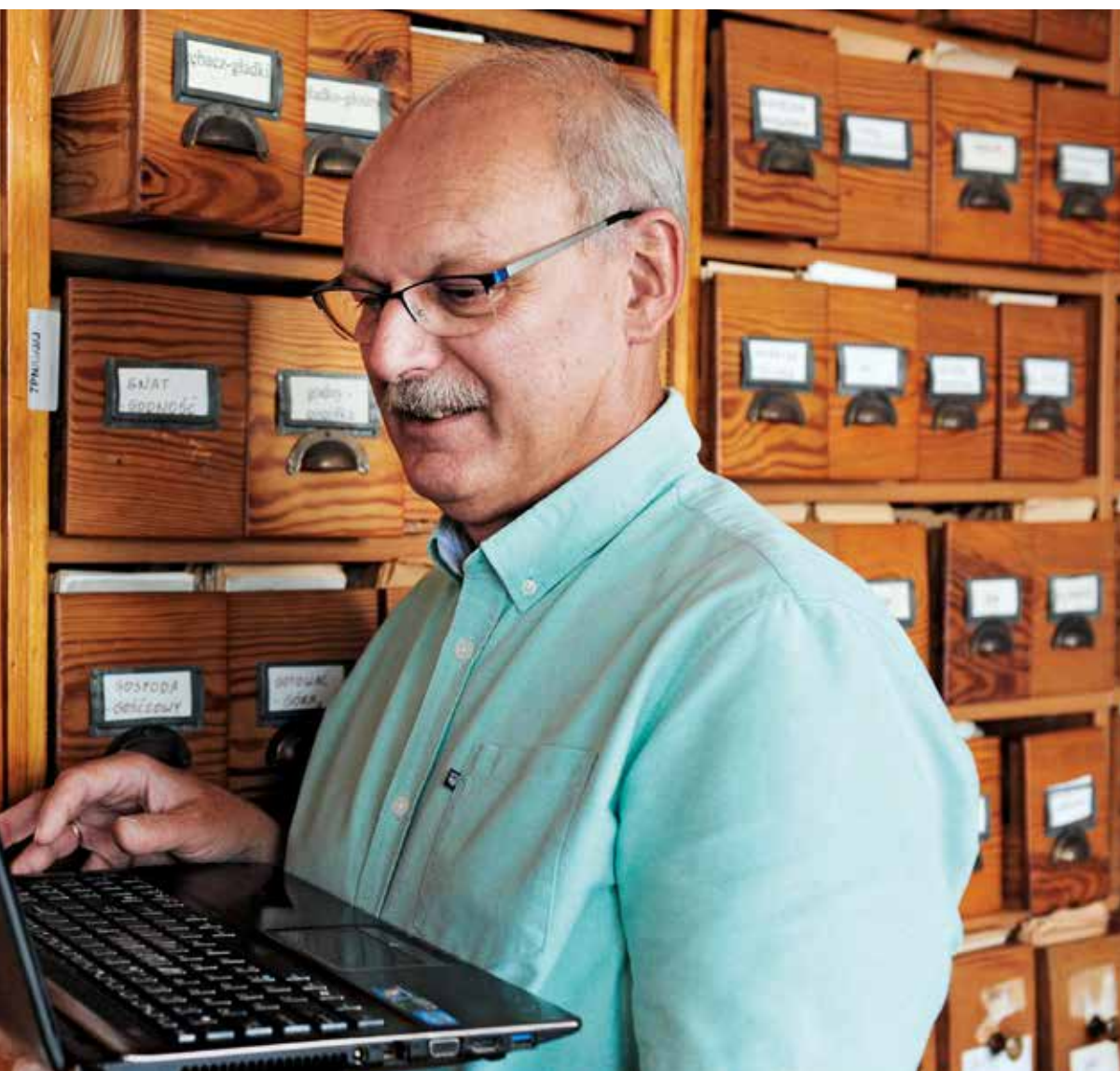
Prof. Włodzimierz Gruszczyński
is a linguist. He heads the Section for the Polish Language of the 17th and 18th Centuries at the Institute of the Polish Language, Polish Academy of Sciences. He teaches classes in Polish linguistic usage at the SWPS University of the Humanities and Social Sciences in Warsaw. He has published numerous books.

wlodekiewa@poczta.onet.pl

specific addressee. We might say it had already been "searched": there were people who did the job of to-day's search engines. Do you know how many times the whole Bible was published in a Polish translation recognized by the Vatican in the 17th century, in the times when the Jesuits and Catholic Church reigned supreme in Polish culture? The answer is: zero. There were only lectionaries, which contained only what priests needed to say mass. After the experiences of the Reformation, it was thought safer not to give the whole Bible to the nation, because the nation – Heaven forbid! – might actually read and understand it. After the publication of Wujek's translation of the Bible in 1599, the next edition, heavily modified by the Jesuits, was not published until 1732. Nearly 140 years later. Back in communist-era Poland, it was easiest to buy a Bible at the British Bible Society on Nowy Świat in Warsaw: few people realized that this was actually the Gdańsk Bible, a protestant translation dating from the 17th century.

> Feel horrified to see young people with their noses glued to their smartphones? The same worry was felt by parents in the 15th-16th centuries, seeing youngsters with their noses stuck in books.

What do you think about how the editors of Oxford Dictionaries picked the "face with tears of joy" emoji as their "word of the year" for 2015?
From a formal point of view I would protest, because an emoji is not a word, but rather a means of expressing emotion.

Still, it performs the function of a word.
More that of a punctuation mark. The traditional exclamation point or question mark also express emotions. Our repertoire of graphical marks has simply expanded. In my view that is something entirely natural, if one looks at the history of writing. In Rome the first inscriptions were written or etched letter by letter, without any spaces between words. There were exactly as many letters on the stone as would fit, without any punctuation marks. With time, dots appeared to mark the end of a word, later segmentation into sentences, and finally question marks and exclamation points arrived. Now we have emoji.

Yes, emotions are easier to show than to describe.
Yes, and it's also fun. For my generation this is still something new, but I would not get upset about it. I can see it situated on a timeline as the next logical change, something that seems completely obvious to the young generation. I am somewhat surprised by something else: young people today, who are said to live in an epoch of images, an epoch of secondary illiteracy, nevertheless prefer to send text messages than to leave a voicemail, write an email, or make a call on Skype. Why is it easier for them to write than to speak? I have a hypothetical answer: you can write to many people simultaneously. And you can also do other things at the same time, such as listening to the radio or watching a film. It interferes less with other activities.

Is it not also the case that text is easier to hide behind?
Yes, but it has one characteristic that I dislike. *Verba volant, scripta manent* – words are fleeting, writing remains.

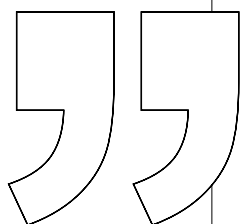**The kind of things some people write on Facebook or Twitter...**
I think a number of people must have lost their jobs when the companies found out what their employees really thought about them.

The written word allegedly requires a moment of reflection.
That depends what we mean by the word "written." The traditional division has always been between written vs. spoken language. This is not just concern the physical medium, but also that written texts are traditionally formulated differently than spoken ones; the former were always more official. But ever since electronic communicators became widespread, however, the idea has arisen that a text written on a keyboard, although physically a written text, nevertheless is similar to spoken language in terms of structure, style, and other characteristics. The phrases used for greetings and farewells, for instance, are lifted straight from spoken language. Usually the authors do not think much about comma use, spelling, or typos. Many people are not bothered by that. In spontaneous communication people give each other a lot of leeway. So this is a completely new, intermediate variety of language.

Does it need a new name?
The term "spoken-written language" has started to be used in a lot of analyses. A kind of mixed form. I'll again refer back to the 17th century: if we look at just about any manuscript from the period, such as a particular letter written by Krzysztof Opaliński, who had studied at the academy in Poznań, the text messages written by young people today simply pale in comparison. The orthographic mistakes made in that letter are so bad as to make it uninterpretable. I used to cringe to see young people write *sie* instead of *się* in Polish, until I saw how Opaliński used no diacritical marks whatsoever in that whole manuscript.

So everyone just tries to save themselves effort?

In a certain sense. And they think: dear reader, you are not stupid, so you know it's supposed to be a letter "ę" here, with a diacritic, not just an "e."

So we don't need to go railing about how this is a lack of respect for language, because the language will manage to defend itself?

Or because someone will defend it. Even in the backward 17th century, there were also people who fought for it. Purists, people concerned for linguistic refinement, do sometimes go overboard. The blanket deniers do so, too. But thanks to the fact that we have both, there is a certain kind of balance. Aside from that, one has to recall that we have different varieties of language for different circumstances. We cannot apply the same standards that are adopted for books printed by renowned publishing houses to a text message written by a teenager. And there are also spontaneity and linguistic wit to consider. If we leave a scribbled note for a housemate, is someone really going to consider if the commas are in the right place? We dress up what we say differently for different occasions. That was and is inevitable.

But language does attest to its user.

That's true, but we choose this clothing also in view of our company. I talk one way in a yachting club to the people I sail with, differently at a social gathering, differently when someone interviews me on radio or TV. And differently in the lecture hall. Of course I don't try to act like a hipster, but I do talk to students in a less stand-offish way nowadays than I did 20-30 years ago.

You are fighting for language in the times of the Internet, by preparing the Electronic Dictionary of 17th and 18th Century Polish, together with your team. Please excuse the question, but I'll ask it point-blank: what is it for?

That is a very apt, but difficult question. I'll answer it by asking another question: what is any dictionary really for? Dictionaries are a very old genre. The first dictionaries in Poland were bilingual, for learning Latin: either passively, to be able to read a Latin text and understand it, or actively, to speak and write Latin. There were also thesauri, meant to give access to the whole richness of Latin. First monolingual ones, then ones with added equivalents in various European languages. The word *calepin*, *calepinus* functioned in nearly all languages in 16th century Europe: it was actually the name of a 15th century monk who created one such a thesaurus for Latin, and became the name of the kind of book. It was thought that every manor library should have a *calepin*, even if it happened to be written by someone else. There were two objectives: teaching and facilitating access to the works amassed by another nation, civilization or culture. That is the case for our dictionary as well.

If an ordinary modern Pole opens up the scan of a book published in 1630 in a digital library, he will immediately have trouble with reading the Gothic type. And even if he can make sense of that, sometimes he will have trouble understanding the words. Sometimes he will not be familiar with certain expressions, but even worse, there will be words he thinks he understands, but in fact they once carried different connotations. A simple example, the word *gęba*, which today is a quite pejorative term (akin to "mug" or "gob" in English), was then simply a normal and neutral word for "face." Or consider *ślimak* (meaning "snail"). In modern Polish the word symbolizes something slow, but in the 17th century it meant something disgusting. Today we know that snails are mollusks, but they were once classified as amphibians or reptiles. Crawling animals were generally considered disgusting, and reptiles were essentially vermin. The snail was then in the quite disgusting company of "vermin". Or another example:

the word *bazyliszek* ("basilisk"), which denoted a real animal for 17th century Poles, one that actually had a deadly gaze.

Everything changes as time goes by and certain subtle traits of words shift, their pragmatic or emotional load, their evaluative content. If someone has not analyzed a large enough number of texts in this dead language – and 17th century Polish is indeed a dead language – it is completely foreign to him or her. We are caught up by various pitfalls when reading such writers as Potocki, Morsztyn, Pasek. If we are dealing with a text without critical annotations, without footnotes or glossaries, we are at a loss and sometimes understand things quite the wrong way. A great example can be found in certain epigrams by Potocki that are included in school textbooks. Some of them are actually obscene, but the textbook authors failed to realize it. How many Polish teachers know the origins

Screenshot from the Electronic Dictionary of 17th and 18th Century Polish.

of the adjective *kiepski* ("lousy")? In fact it derives from the noun "*kiep*", today associated with a cigarette butt, but in the 17th century it was an indecent term for the female reproductive organs. Later it became an epithet for a man with feminine qualities, a loser, joker. Nowadays when I hear a priest talking about *kiepskie zachowania* ("lousy behavior")…

All of this was a lengthy answer to the question of what this dictionary is for.

In short: to understand the times someone is describing, one has to understand the language they used.

Yes. But this is definitely not a dictionary for the average Polish man-in-the-street. Although it is still more easily readable than the *Dictionary of 16th Century Polish*, for instance. It is relatively easy for laymen to use. But since not many people nowadays read 17th century poetry before going to sleep, the dictionary is mainly addressed to publishers of annotated editions of old Polish literature who want to see in what contexts the word appeared in the works of other poets and writers, how others used it, and what remarks we, the authors of the dictionary, have to offer about the given word. Most importantly, we do not have any monopoly on saying how these words should be understood; these are just our hypotheses. After reading through many contexts we try to figure out how the word may have been understood, how it was used, we list all the phonetic forms we are aware of, all the inflections and variants, so that researchers in other fields (historians, literary specialists) can utilize this information. Our colleagues from the Institute for Literary Research and the Institute of History are especially pleased that the dictionary is being written.

How many entries does it have at this point?

Nearly 28,000, in varying degrees of advancement, and there will be more than 80,000.

How many people are on the team?

At this point 7, but in the best times we had a few more. Today everyone chooses the entries they find to their tastes. Some take weeks to work on, because they require consultations with someone knowledgeable on the issue: perhaps a biologist, orientalist, or Turkologist.

What kind of work do you handle?

I most often double-check what my colleagues write, because I have to approve the entries. That is the most ungrateful task, because I'm fixing errors and inconsistencies. And also structural problems, because our editorial guidelines are still a work in progress, even though we have been working on the dictionary for years. We are constantly adapting them to the material and to other dictionaries, so that ours will be comparable to them.

How long has the dictionary been under creation?

The card file we have here was created in 1954. Two years ago we should have celebrated the 60th anniversary of work on it. We presented the online prototype at the Science Festival in 2005, with the young, then up-and-coming programmer Mateusz Żółtak. Then came the first stage of dictionary work: "porting" the entries from the paper version to the electronic version.

Which means?

Transferring the entries written on a computer as ordinary text to a database specially developed for our purposes. At the same time, changes needed to be made in the dictionary's structure. For instance, we got rid of the cross-referenced entries. This can be explained most simply with a modern example. If I'm writing a Polish-English dictionary which English speakers are also meant to be able to use, I cannot just have one entry headed by the word *dech* ("breath"), because such users may come across the genitive form *tchu* in a text and will not know that they need to look for it under the nominative form *dech*. So there has to be a cross-reference: *tchu*, see: *dech*. In our dictionary all the variants of a word were treated this way, such as *apteka* ("pharmacy") and the variant forms *aptyka – hapteka – japtyka – apoteka*. In the printed version we had to decide which of those variants to treat as the main headword, with the others being cross-referenced. There were two huge boxes of cross-referenced entries left behind 50 years ago for those who would one day be writing the entries under the letters *h*, *s*, *z*, etc.. It was out of control. Today there is no problem in the electronic dictionary, because every variant essentially has the status of a cross-reference. It makes no difference which headword is typed into the search engine, because they all point to the same main entry. The problem of cross-referencing ceased to exist.

Is the end somewhere in sight?

No. I definitely will not live to see the completion. Not because we are working slowly, but because there are too few of us and we also have other jobs.

Which 17th century word are you most fond of?

Babadziafer. Firstly, because it sounds nice, even though it means "prison, place of torture." Secondly, because we found it attested only in a single text, which is truly extraordinary. Thirdly, because for a while our website had a URL that was too complex to remember, but I could just Google the word *babadziafer* and go straight there. So I know that thanks to this word one can easily find us online.

INTERVIEW BY **KATARZYNA CZARNECKA**
PHOTOGRAPHY BY **JAKUB OSTAŁOWSKI**