

Scientific Approach to Etymology of Surnames

by Alexander Beider
Paris

The aim of this paper is to show a possible approach in the domain of etymology of surnames for which the label „scientific“ would be appropriate. It discusses the reasons of the low status of that branch of onomastics in scholarly circles. A scientific approach asks to provide a conjecture that would be the most corroborated by various factors among which are: answers, correlated with the etymon and between them, provided to other main questions of the etymological research; results of the statistical analysis of surnames of other families belonging to the same ethno-cultural group; knowledge concerning the general historical and linguistic background of the time of the surname adoption; finding of the family genealogical data concerning the time of the surname adoption; and existence of the testimony concerning the surname adoption.

Introduction

The modern Webster's dictionary of English provides two meanings for the term "etymology":

1. The history of a linguistic form (as a word) shown by tracing its development since its earliest recorded occurrence in the language where it is found, by tracing its transmission from one language to another, by analyzing it into its component parts, by identifying its cognates in other languages, or by tracing it and its cognates to a common ancestral form in an ancestral language.
2. A branch of linguistics concerned with etymologies.

Names 53:1&2 (March & June 2005):79–126

ISSN:0027-7738

Copyright 2005 by The American Name Society

Below, only the first meaning will be taken into consideration. One can say that the etymology of a word is its linguistic history, including its origins and derivation. This general definition can be directly applied to common nouns, adjectives, verbs, and all other words different from proper names. It can also be used in the context of personal (given) names.¹ It is, however, only partly appropriate for surnames (family names).² Indeed, numerous surnames are not transmitted from one language to another, it is impossible to find any cognates in other languages, and in a large number of cases the ancestral form is identical to the form used in our days. As a result, by withdrawing from the general definition non-applicable parts, one can consider that the etymology of a surname consists simply of suggesting its source word (*etymon*), that is, the word from which it is directly derived.

Taken in this narrow sense, the study of the etymology of surnames of particular ethno-cultural group³ is characterized by a feature that is very unusual in scientific matters: in principle, definitive knowledge is possible. This is due to two obvious factors:

1. The surname of each family within the group was constructed from a specific source word (or expression) that was undoubtedly known either to its first bearer or to the person who created the surname (for example, a local authority). This factor does not directly apply to those families whose surnames resulted from the gradual transformation of nicknames into surnames. Even for these families, however, definitive knowledge is still possible: the origin of the nickname was known to the person who used it for the first time.

2. The number of different surnames used within the group is finite.

Consequently, if we (a) compile the exhaustive list of surnames borne by representatives of an ethno-cultural group during a specific time period; (b) find the source words for every surname from that list, for that group the task of the

etymological research will be *totally* accomplished. The creation of the list with two columns, surnames and their respective source words, will preclude any additional study.

This formal possibility of solving *all* its problems *definitively*, closing this way the whole domain of study, puts the etymology of surnames in a very special position that can be psychologically judged from different points of view. Some people would be excited by the idea of definitive answer to stated questions which is generally associated only with formal sciences, such as mathematics and logics. The reaction of others would be skeptical. The arguments can vary, depending on the field of scholarly activity in which the person is involved. One can, for example, imagine the following argumentation provided by an amateur of formal sciences. If indeed the absolute truth can be reached in principle in both mathematics and etymology of surnames, a mathematician can really *prove* that he⁴ is right, while an onomastician is unable to prove anything. Reliable proof of the accuracy of a particular etymology may lie, would say the same skeptical person, only in the discovery of a document written by the person who created this surname in which the source word is given explicitly. This type of primary evidence is, however, almost non-existent. Any other evidence suggested by the onomastician represents no more than his own opinion resulting from his *belief* in what is true, and beliefs should be beyond any science. As a result, the etymology of surnames is not a scientific matter.

The idea of the creation of a two-column list as the ideal final aim of the project dealing with etymology of surnames, reveals another peculiarity of this domain. In the list in question, every line will be, in principle, totally independent from others. To tell it in a less formal way, a suggestion of the source word for one surname can be true or false independently of the truth or the falsehood of the suggestion concerning another surname. An onomastician who prepares an etymological study of surnames used by one

ethno-cultural group does not suggest one theory, but as many theories as the number of surnames in his collection. This feature makes the search for surname etymologies very distinct from other scholarly domains. Indeed, the specialists in natural sciences or humanities generally try to suggest theories that explain numerous facts. The more general the scope covered by a theory and the larger the number of its factual consequences, the more esteem it gets from the scientific and—due to the high status of the scientific knowledge in our civilization—extra-scientific community. Nothing similar can be said about the etymology of surnames. As was said above, in principle, a scholar provides as many “theories” (suggested source words) as “facts” (surnames) he tried to explain. This feature can also yield skeptical views concerning the possibility of considering this domain scientific. An onomastician could be compared to a collector of stamps: very systematic, of good general culture (history, geography, foreign words), knowledgeable about certain important, in his opinion, characteristics of the elements that compose his collection, often fascinated by it, and willing to share his fascination with other people. A very similar comparison could be made with a lay entomologist. There is nothing wrong with these human activities, but usually they are not usually considered to be scientific.

Another characteristics of the domain of the etymology of surnames consists in the fact that for numerous surnames their source words are self-evident. Indeed, for anyone in the United States and the United Kingdom, the origin of James, Rodgers and Johnson, Smith, Mason and Baker, Long and Snow is trivial. McDonald and McGregor, Fitzgerald and Fitzpatrick are more complicated, but once one knows that *Mc* comes from Gaelic *mac* and *Fitz* from Anglo-Norman French *fi(t)z*, both meaning “son of,” these and numerous other Scottish and Irish surnames with the same prefixes cease immediately to pose any problem. Similarly, common French surnames Martin and Thomas, Dupont (from the bridge) and

Dubois (from the forest), Boulanger (baker) and Meunier (miller), Petit (small) and Leboeuf (the ox) are not an enigma for anyone fluent in French. A small intellectual effort is needed to understand that Lefèvre comes from old French "the smith" and Maréchal is cognate to English Marshall. Numerous examples of the same kind can be easily suggested for various other nations and/or ethnic groups. Numerous also are surnames that are young, assigned during the 19th-20th centuries. For them, often the source words are obvious. Needless to say, the triviality is not typical for scientific matters.

The above considerations shed light onto the history of the etymological research in the domain of surnames. The aforementioned factors are likely to be (at least partly) responsible for the low status of that branch of onomastics in scholarly circles resulting in (1) a large number of popular books and articles compiled in this domain, and (2) a very small number of serious studies, generally relatively recent.⁵

The aim of this paper is to show a possible approach in the domain of etymology of surnames for which the label "scientific" would be appropriate. Throughout this paper, the word "scientific" will mean, as it usually does in many scholarly works, "leading to the discovery of the *objective* truth, the truth that corresponds to facts." Consequently, the focus will be on elements that belong to the domain of objective knowledge. Various psychological factors will be deliberately ignored, as they are necessarily subjective. To explain the etymology of surnames that are products of human thoughts, there is no need to resort to psychology. Quite the opposite: the results of the onomastic analysis can, in principle, be used to explain certain aspects of individual and social psychology.⁶

Most of the principles exposed here are not original. Generally, serious studies in etymologies of surnames make use of them without formulating them explicitly. The main aim of the present text is to cover this deficiency and to

provide a formal outline of methodological principles that could be used as normative.⁷

1. Basic Theoretical Frame

First of all, it needs to be stressed that, formally speaking, any suggested etymology cannot be taken as absolute truth. Intuitively, the discovery of a document written by the person who created the surname in which the source word is given explicitly, seems to represent such kind of definitive proof. Still, in many cases this idea is also no more than an illusion. To understand it, it is sufficient to take into account the principles, basic in historiography, which imply that (a) before relying upon any historical testimony, its authenticity has to be questioned (for example, a document could be compiled by someone who was not directly involved in an event); and (b) the factual information present in a narrative source is not necessarily truthful. Any direct testimony is invariably intentional, lead by its author's motives, among which the desire to document the events as close to the reality as possible is not a unique one; other psychological elements can be of more influence.⁸ Finally, a piece of information, when being documented, can be distorted unintentionally, due to memory failures.

Of course, it would be an exaggeration to consider that the cases are nonexistent when documented evidence is totally reliable. Indeed, it is, for instance, the case of a large number of surnames assigned through an official procedure during the 20th century. Consider two examples. During the first decades after the creation of the USSR, for numerous representatives of various Muslim ethnic groups living within the country without any previous use of hereditary names, had their surnames created by the addition of Russian possessive suffixes *-OB* [ov] and *-IH* [in] to the given names of their fathers. In contemporary France, foreigners acquiring French citizenship receive a paper that shows them the two ways they can acquire, on their own will, a "French" surname: either

translating the literal meaning of their current foreign surname into French, or by choosing an existing French surname that has some phonetic similarities with their current surname. These cases are, however, rare exceptions. They are limited to surnames acquired recently and do not even cover the whole category of names created during the 20th century. For example, people whose ancestors emigrated to US before World War I changed their surnames during the first years that following their emigration, often cannot discover any written document that would explain the way the new surnames were chosen. Documented evidences are almost never found and if they are found, in most cases, they cannot be considered to be totally reliable.

This results in the following fundamental principle: generally speaking, an etymology suggested for any surname is no more than a hypothesis that cannot be formally proven. If etymologies can never be "true," there are cases when one can present arguments that force to reject a proposed etymology as "false." The examples will be given in the following sections.

As any hypothesis, an etymology can be more or less plausible. The text below suggests the way its verisimilitude can be evaluated. No numerical estimation will be suggested: there seems to be no sense introducing any subjective conventional rule that would enable to calculate the verisimilitude of a hypothesis saying, for example, that it is 0.5 and not 0.8. There are, however, less conventional rational criteria that allow to address the problem of verisimilitude in relative terms. The comparisons can be done on two levels. It makes sense to compare either two concurrent hypotheses concerning the same surname, or the hypotheses concerning two different surnames. In both cases, one compares two theories trying to understand the verisimilitude of which one of them is greater.

Before proceeding to the comparison, several research steps should be performed.

- A clear distinction should be made between the direct source words (etymons) and indirect sources. •For any surname under analysis, one should try to find not only its source word, but also answers to several other basic questions of the etymological research.

- For any surnames under analysis, one should try to find genealogical data concerning the family of the first bearer of the surname

- For the ethno-cultural group to which the first bearer of the surname belonged, an analysis of the historical background should be performed focused on the period when the surname was created

- For other surnames assigned during the same period within the same ethno-cultural group, a statistical analysis should be performed.

All these steps are discussed in detail in the following sections. Without the supplementary information obtained during these steps, hypotheses concerning etymons have no objective basis and should be treated as speculative, *ad hoc*.

2. *What*-question and Classification of the Source Words

2.1 The conjectured source words, also called *etymons*, represent the answer to the question *What word or words were the direct basis for the surname?* In the text that follows, this will be referred to as *What*-question. One who tries to suggest an objective answer to this question needs to make this answer geographically determined: it should concern only one specific continuous area. Various surnames borne by representatives of the same ethno-cultural group living in that area, can be classified into several categories depending on their source words which can be:

- Words (mainly common nouns and adjectives) and expressions taken from the general vocabulary of the language used by this group at the moment when the surname in question arose for the first time in the same area. This language generally belongs to one of the three following cases:

either it is spoken in the vernacular, or it is official, used by the country administration, or it is purely cultural, as Latin and Greek for various European Christian people, Arab for Muslims, and Hebrew and Aramaic for Jews;

- Given names (first names, forenames) used within the same group at the time of surname adoption;
- Placenames (toponyms);
- Surnames used already within the same ethno-cultural group;
- Surnames borrowed from the representatives of another group living inside the same area;
- Surnames brought to that area by migrants from other geographic regions.

The first three categories are the most universal, found in numerous cultures and peculiar, for example, for all of Europe. Very often surnames are derived from these sources not directly but via personal nicknames that become fixed as hereditary. In several ethno-cultural groups - such as Turkish, Ashkenazic Jewish, and Danish - surname adoption was forced by officially promulgated laws and surnames can often be really dissociated from nicknames. For numerous other people—English, French, German, Italian, Russian, and Polish are among the examples—a very large number of surnames result, however, from nicknames. In this case, one could say that surnames are no more than a subset of nicknames, namely nicknames that changed their status and became hereditary. Non-hereditary nicknames necessarily have their own etymons that can in turn be assigned to one of the categories of the above classification. Indeed, most nicknames are drawn from words and expressions taken from general vocabulary and placenames. Sometimes they can also be based on given names. Some of them can, in principle, be also derived from other nicknames, or even from surnames. In a very small number of cases, nicknames can be derived from surnames or nicknames either brought by migrants, or used by the representatives of another ethno-cultural groups. As a result,

even if one should formally acknowledge the possibility that the source word for a surname can be a nickname, an onomastician will never be totally satisfied with this answer trying to discover the source word for the nickname itself. In other words, when doing etymological research, nicknames and surnames can be treated together. For these reasons, nicknames are withdrawn from the analysis that follows. For our specific methodological considerations, they are of no interest. If a surname is actually derived from a nickname, the source word for the nickname will be conventionally considered in the text below to be the etymon for that surname.

Examples of the last three categories are less frequent. All of them have the same common feature: here, in contrast to first three categories, other surnames are the source words. In the fourth category, both the new name and its etymon, an old name, belong to the same ethno-cultural group and correspond to the same language. Two major subsets can be discerned. The first one covers surname changes when a Family X borrows its new hereditary appellation from Family Y belonging to the same group. In this case, the etymon (surname borne by Family Y) and the new name (borne by Family X) are phonetically and graphically identical. The second subset encompasses the cases when both the new and the old name (its etymon) belong to the same family. It covers (a) intentional name changes that take place without any migration and in which the new surname is derived from the old one by dropping, adding or substituting certain letters; (b) unintentional distortions occurring in surnames.

The fifth category encompasses, almost without exception, only surnames borrowed by representatives of a minority from the corpus of surnames used by the majority. Among the examples are those Jews and Gypsies who have adopted, in various countries, surnames borne by the local population, numerous immigrants who took standard English, Scottish and Irish surnames upon their arrival in Northern

America. The situation with the names of African-Americans is similar: their first black bearers generally received typically British surnames of the owners of plantations at which they worked. The last example shows that in this context, it is more appropriate to measure the difference between “minority” and “majority” not numerically, but in politico-economical terms. Another example in which the validity of that statement is even more striking concerns Livonia and Courland, that is the area that currently correspond to western Latvia. Here, during the 19th century numerous Latvian-speaking peasants received German names from German-speaking officials, the representatives of the politically and economically dominant group of these provinces of the Russian Empire.

2.2 The last, sixth, category encompasses surnames borne by immigrants who retained in the new area phonetic and/or graphic elements of the surnames they used in the old. Excluded are, however, surnames the modification of which resulted in forms of one of the two following sorts:

- Normal words of the local language. In this case, the words in question - and not the surnames used in the old country - will be conventionally considered to be their etymons. As a result, these surnames are assigned to the first category;
- Surnames used by local population. In this case, these surnames - and not those used in the old country—will be considered to be their etymons. This brings the appellations under analysis into the fifth category.

To illustrate this definition, consider several appellations of German immigrants to US. To the last, sixth, category belong, for example, Dreiser (both phonetics and spelling were retained), Steinbeck (only the spelling was retained, while the pronunciation changed), Kline (the pronunciation of Klein kept), Eisenhower [from Eisenhauer] and Steinway [from Steinweg] (one of two parts remained German), as well as Hild [from Hildebrand]. On the other hand, forms resulting from direct translation (calques) such as Black [from Schwarz], Eagle [from Adler], Miller [from

Müller], and Field [from Feld] belong to the first category.⁹ The forms in which not the semantics of the original German surnames, but their phonetics was taken into account, also correspond to the same category: Grove from Graf [earl] and Cashdollar from a dialectal pronunciation of Kirchthaler [person from (a village of) Kirchthal]. The etymons of the above German-American surnames are English words *black*, *eagle*, *mill*, *field*, *grove*, and *cash-dollar*, respectively. On the other hand, German-American surnames like Campbell, Stewart and Webster, independently of whether they possess or not possess any phonetic similarity with the surnames used in Germany, will be placed into the fifth category. Their etymons are identical British-American surnames.

The surnames from the sixth category differ from all the other ones. Their sources are foreign for the area under analysis. In contrast to the names from the first four categories which were created within the same geographic area from local sources, and those from the fifth one which were borrowed from local sources as well. For methodological reasons, a name brought from another country will be conventionally treated during the etymological analysis as a new name in comparison to the old one, used in the country of origin.

The validity of this principle is evident in cases when one or several elements of the old name are changed in the new country, say US. For example, Italian Cristiano and De Franco transformed to American-Italian Christiano and De Frank, respectively. Or Polish Kowalski and Jabłoński turned to Koval and Jablons. In some cases, only diacritical signs are dropped as in Lefevre from the original French Lefèvre and Benes from Czech Beneš. The change of the alphabet necessarily produces a new graphic form. As a result, the surnames of people who migrated from Russia, Serbia and Bulgaria to Northern America and substituted there the Roman alphabet names for those originally written in Cyrillic letters, are new in comparison to their old, European, names,

even if they are pronounced in the same way. Similarly, modern Israeli names that are written in the official documents only in Hebrew are necessarily different from names used by the same families before the emigration, originally written in Latin, Cyrillic and Arab letters.¹⁰ In certain families, the spelling of their surname was kept in the new country, but its pronunciation changed following the reading rules of the official language of that country. Typical examples are American-German Sommer and Zimmermann. In Germany, their initial letters were pronounced /z/ (as in English “zero”) and /ts/, respectively. In the United States, the phonetic value of these letters was changed: in the first one it is now often pronounced as /s/ (as in “sea”) and in the second one it is pronounced /z/.

There are also cases when the name in two countries, that of departure and that of arrival, is both written and pronounced in the same way. It is a very common situation if in both areas the same vernacular language is spoken. This feature characterizes, for example, surnames brought from British Isles to Northern America or Australia, and those that came from France to the Canadian province of Quebec. Even in these cases, however, this time purely conventionally, when discussing their etymologies, the two names will be treated in this paper as different. For example, the common noun *bæcestre*, meaning baker in old English, represents the etymon for the surname Baxter in England. This English surname (and not the occupational common noun) will be considered to be the etymon for the American-English or Australian-English surname Baxter. The main reason for this distinction comes from the following principle: the etymology of any surname should be discussed specifying explicitly the ethno-cultural group to which it belongs.¹¹ In the above example, English, American-English and Australian-English are distinct groups.

3. *Why*-question or Reasons for the Choice of Source Words

3.1 The purely linguistic definition of etymology presented at the very beginning of this paper is rich enough to stimulate an etymological research concerning words of general vocabulary. In the context of surnames, as it was already discussed, several elements of this definition are not applicable and those that remain would restrict the task of onomastician to the discovery of source words. This definition is generally felt as too narrow and it is not a surprise that rare are studies in this domain whose authors limit themselves to the search of direct etymons and explanations of differences, if any, between surnames and their etymons. Usually, scholars complement their analysis by another, additional, question, namely *Why has the source word been chosen to construct the surname?* This question (below referred to as *Why-question*), forces a researcher to exit from the linguistic context to enter another, supplementary, domain, a richer one, prompting for more imagination and therefore a more fascinating one, but invariably more subjective and consequently more easily exposed to accusations of not being scientific.¹² Needless to say, in many cases answers to the questions of this kind would be purely hypothetical and no absolute truth could be reached.

If one does not want to deviate from a scientific approach, a very specific, narrow, understanding of *Why-question* is imposed: one has to search only for direct reasons that can, in principle, be considered as objective. Fortunately, for a large number of surnames such reasons can indeed be found. On the other hand, certain questions that deal with deeply psychological reasons are outside of the scientific approach in onomastics. Among the examples are:

Why did one individual adopt a surname related to his place of origin, while another, during the same period and in the same place, choose a surname drawn from the given name of his father?

Why did one individual construct the surname from the common noun that designated his occupation, while another,

whose occupation was identical drew it from the common noun that designated the main tool used in this activity?

If for one specific surname the answers to such questions are totally subjective, statistical analysis of the whole corpus of names adopted within the same ethno-cultural group can, nevertheless, reveal some characteristics specific to that group.¹³ Even if this procedure gives no answer to *Why*-question for particular instances, it provides results that allow to observe psychological features that are common to numerous representatives of the same group. In this way, these results belong to the domain of objective knowledge.

3.2 In a number of cases, the creation of a new surname from its source word is due to the existence of additional, third, word, and without taking it into account, the answer to *Why*-question would be incorrect. Several examples of this kind were already presented in section 2 when speaking about surname changes made by immigrants. For instance, German-American surname Black is drawn from the English adjective *black* (its etymon) because the surname previously used in Germany was Schwarz (meaning *black* in German). Here, it is appropriate to speak about the existence of the semantic association between the new and the old surnames. Similarly, Polish-American Miller is derived from the English common noun *mill* because the original surname used in Poland—such as Młynarski, Młynarczyk, and Młynarz—had the root meaning *mill* in Polish. When languages used in two countries are related, the association can be often both semantic and phonetic, as in such German-English pairs as Feld > Field, Blum > Bloom, Stein > Stone. In other surnames, the association is purely phonetic. This phenomenon can be observed in the derivation of German-American Beam. Its etymon is a common English noun *beam*, but the choice of this source word was guided by the fact that the original surname used in Germany was Böhm (from *Böhme*, one from Bohemia), that has the same consonants as *beam*.

A phonetic association is also found in a number of surnames borrowed from the representatives of another group living inside the same area. For example, common Jewish-American surnames Harris and Davis (borrowed from American Christians of English stock) were generally chosen because the surnames used by the same families in Eastern Europe prior to emigration were (1) for Harris: various forms based on Yiddish male given names Hersh and Hirsh such as Russian *Гершкович* [Gershkovich], *Гершезон* [Gershenzon] and *Гиршман* [Girshman], Polish *Herszkowicz*, *Herszlikowicz* and *Herszman*, German *Hirsch*, *Hirschmann* and *Hersch*, and (2) for Davis: surnames derived from the given name David, such as Russian *Давидович* [Davidovich] and its Polish equivalent *Dawidowicz*.

All of the above examples are directly related to migrations between countries. A very special case of semantic association is covered by calques, that is, surnames resulting from a translation—made without any migration—of surnames from one language to another. To that category belong Finnish calques of former surnames drawn from Swedish words. These new surnames, inspired by Finnish national feeling and fashion effects, appeared during the 19th-20th centuries in Finland. Similar situation characterizes Latvia where during 1930s numerous surnames were translated from German to Latvian. During the 15th-16th century, a number of German and Dutch surnames were translated into Latin following the fashion tendencies of the Renaissance humanism. Among the examples are *Avenarius* from *Habermann* (from the root meaning *oats*), *Faber* from *Schmidt* and *Smit* (smith), *Minor* from *Klein* (small), *Molitor* from *Müller* and *Möller* (miller), *Piscator* from *Fischer* (fisherman). A number of Russian surnames assigned during the 18th-19th centuries to students of the Orthodox theological schools represent Greek or Latin calques of the original Russian surnames: *Артобелевский* [Artobolevskiy] from *Хлебников* [Khlebnikov] (in both cases, the root means *baker*), *Струтинский*

[Strutinskiy] from *Воробьёв* [Vorob'ev] (*sparrow*), *Лепорский* [Leporskiy] from *Зайцев* [Zaytsev] (*hare*), *Ансеров* [Anserov] from *Гусев* [Gusev] (*goose*). In eastern Poland, during 1820 Polish Christian clerks created a series of Jewish surnames representing Polish calques of original German surnames assigned some twenty years before by Austrian Christian officials: *Białezkło* from *Weissglas* (white glass), *Czerwonykamień* from *Rothstein* (red stone), *Jabłonkowiedrzewo* from *Apfelbaum* (apple tree), *Niebieskafarba* from *Himmelfarb* (sky-blue paint), *Pięknawieś* from *Schöndorf* (beautiful village), *Różowagóra* from *Rosenberg* (rose mountain) etc.

For surnames created as calques, the etymons—that is, the answers to *What*-question—belong to the first category, that of general vocabulary words. Semantic association with previously used surnames represents the generalized objective answer to *Why*-question.

In all examples considered above in this section, the association was made to previously used surnames. New surnames making (phonetic) association to given names do exist too, but they are significantly less common.

3.3 In cases when no additional association is being made, the consideration of the semantic field to which the source word belongs can also often provide direct clues as to why it was chosen. The following classes are shared by numerous ethno-cultural groups:

- Occupational surnames in which the words chosen to form surnames indicate the occupations of the first bearers or of one of their parents;
- Surnames drawn from words that designate physical or moral qualities of the first bearer, his family situation, social status, belonging to special castes, topographic characteristics of the place of residence or other traits of his house (often via nicknames);
- Surnames derived from the given names of the first bearers, their parents, grandparents, or spouses;
- Those based on

masculine names are called patronymics, those created from feminine names are called metronymics;

- Surnames drawn from names of places and regions of origin or residence of the members of families of the first bearers.

The above classification allows to provide objective answers to *Why*-question for surnames that fall into these classes: the source words provide direct references to characteristics of the families of the first bearers. For example, such appellations as English Smith, German Schmidt, French Lefèvre, Spanish Herrero, Portuguese Ferreiro, Italian Fabbri, Czech Kovář, Polish Kowal, and Russian Кузнецов [Kuznetsov] were acquired because their first bearers or their fathers worked as smiths.¹⁴ Still, certain subjective nuances can be discerned even for these four classes, especially for the second one. Firstly, a surname (or a nickname) can be drawn from a word designating some human quality either in a direct way, or it can also be applied ironically, that is, because the actual characteristics of the individual in question was quite the opposite. Secondly, the meaning of the source words can be ambiguous. For example, a surname drawn from a word meaning *white*, is it pointing out to the complexion, the hair color, or, perhaps, something else?

The subjective elements are significantly stronger in cases when the references provided by etymons are indirect. This covers surnames derived from common nouns that designate various representatives of flora and fauna, metals, minerals and other natural objects, food, parts of body, miscellaneous manufactured objects, etc. For most cultures appellations of this kind were initially created as nicknames that later became hereditary. In numerous cases they make indirect references to some physical or moral characteristics of their first bearers. A number of other names point to occupations. This is often the case when source words correspond either to house utensils, working tools, cloths and other human objects, food, etc. These surnames reveal

manufacturers of these elements or dealers in. In onomastics, they are called metonymic occupational.

3.4 The reasons for the borrowing of surnames from the representatives of another group living inside the same area are socio-psychological. They are often due to the assimilation tendency of the minority to the majority. This tendency can be either (1) forced, imposed by the representatives of the majority, or (2) natural, when the idea of borrowing comes from the future bearers. The former situation is objective, in a sense that it concerns not a few individuals, but the whole group of people for whom surnames are assigned. The latter situation is more subjective since the choice of his own surname made by any particular person is not constrained and, in principle, every case is specific. There are, however, periods when numerous members of the same ethno-cultural group borrow family names of the dominant group by their own will. This occurs when the assimilative tendencies are quite general and their objective existence can be shown by the statistical analysis. Surnames taken by Hungarian Jews during the second half of the 19th century and the first part of the 20th century can serve as illustration. In numerous cases, the newly adopted surnames were neither related to previous surnames (mainly drawn from German words), nor to any characteristics (such as occupations, physical or moral peculiarities, given names, places of origin or residence) of their Jewish bearers, but resulted from the direct borrowing of surnames commonly used by Hungarian Christians. This phenomenon was one particular manifestation of the general nationalistic climate that characterized Hungary during that period.

General assimilative tendencies can also explain the reasons of some borrowings inside the same ethno-cultural group (a subset of the fourth category of etymons). American Jewish surnames can illustrate this statement. At the beginning of the 20th century, a number of new immigrants from Eastern Europe abandoned their surnames in favor of other Jewish surnames, commonly used already in US: Cohen,

Epstein, Goldman, Rosenberg, etc. These new appellations were judged typically "American" and often they replaced those that included Slavic elements.

Unforced borrowings of surnames used by other families of the same group that are influenced by the will of assimilation represent one particular case of fashion phenomena that reveal a specific peculiarity of social psychology, that of imitation of features of other persons. An example of fashionable name changes can be taken from the Polish history. In that country, numerous peasants added to their original surnames the suffix *-ski* often obtaining this way surnames used already by nobility. To the same type also belong surnames (from the same or another ethno-cultural group) chosen because they have famous bearers. Among the examples are the surnames Lincoln and Washington chosen by African-American families.

3.5 For secondary surnames resulting from modifications, deliberate or unintentional, of previously used surnames (another subset of the fourth category of etymons), *Why*-question makes no sense: the former (primary) surnames are appellations that were not newly chosen, but preexisting.

In the case of surnames that are due to migrations between countries (the sixth category of etymons), this question seems, from the first sight, to be trivial. Indeed, according to the conventional definition introduced in section 2.2, a new surname appears automatically when an old surname is brought to a new country, even if no phonetic and graphic differences can be discerned between the two appellations. Only one specific condition is needed: in the new country, the practice of bearing surnames should be established as in absence of this feature, an immigrant can acquire no surname at all. A more attentive consideration will show, however, that even with surnames resulting from migrations the *Why*-question makes certain sense. In this specific context, it could be reformulated in the following way: *Why were certain phonetic and/or graphic elements of the surname*

that was used in the old country kept in the appellation that was retained in the new country? Legal, linguistic and/or psychological factors can be evoked here. First of all, one can formally imagine a legal situation in which an immigrant has no choice: he has to keep the surname borne in the country of origin. Secondly, if the surname used in the former country does not sound foreign in the new one and it is easily pronounceable here, there is no evident need to change it. Of course, one can imagine individuals that could, in principle, wish to start a “new life” in a new country with a new name. These and other totally subjective psychological motives should be necessarily beyond the scope of the etymological analysis. Thirdly, if the old surname does sound foreign, very common is the desire to change it withdrawing elements that could reveal the geographic and/or ethnic origins of its bearers. In this psychological context, an onomastician willing to restrict his work to the objective knowledge can proceed to the statistical analysis in order to discover the general trend peculiar to that group.¹⁵ Other questions, concerning not the group as a whole, but individuals that compose it, are clearly beyond the scope of (scientific) onomastics. Among the examples are:

- Why in countries where there is a very strong tendency for assimilation, do certain individuals keep the surnames that sound totally foreign?
- Why among the immigrants in search of assimilation, do certain decide to translate their former surnames, others prefer to find local words or surnames close to their old surnames or having nothing to do with them, the third ones modify only one or several elements in them (for example, withdrawing suffixes or changing letter combinations of evident foreign origin)?¹⁶
- Why in certain immigrant families bearing surnames spelled exactly as in their country of origin, is the pronunciation kept as well, while other families start to pronounce it according to the rules proper to the official language of their new country?

3.6 Table 1 summarizes the information presented above. It gives the typology of surnames according to the generalized answers to *What-* and *Why-*questions.

Table 1. Typology of surnames

Surname type	Etymon (<i>What</i>)	Reason for the choice of the etymon (<i>Why</i>)
1.1 Occupational names	Words related to occupations	This occupation was that of the first bearer or some other members of his family
1.2 Surnames designating qualities	Words that designate qualities, family situation, social status	This quality was that of the first member or some other member of his family
1.3 Calques of previous names	Words representing the translation of previous surnames used within the same ethno-cultural group	Semantic association to the previously used surname
1.4.1 Immigrant names created semantically	Words representing the translation to the local language of previous surnames used by immigrants	Semantic association to the previously used surname
1.4.2 Immigrant names created phonetically	Words from the local language phonetically close to previous surnames used by immigrants	Phonetic association to the previously used surname
1.5 Artificial surnames	General vocabulary words that have no relationship to any characteristics of the first bearer	(<i>See the text below the table</i>)
2.1 Patronymic surnames	Male given names	This given name was borne by some male member of the family of the first bearer

Scientific Etymology • 101

2.2 Metronymic surnames	Female given names	This given name was borne by some female member of the family of the first bearer
3. Surnames drawn from placenames	Placenames (toponyms)	This place was that of the origin or the residence of some member of the family of the first bearer
4.1 Secondary surnames (intentional or unintentional changes)	(Primary) surnames used in the same family	Not applicable
4.2 Surnames borrowed internally	Surnames borne by other families of the same group (common names, names used by the dominant social group, names with famous bearers)	<i>Factors of social psychology (assimilation, desire to have an "illustrious" name etc.)</i>
5.1.1 Forced borrowing	Surnames borne by families of another, dominant, group	Forced assimilation
5.1.2 Natural borrowing	Surnames borne by families of another, dominant, group	<i>Factors of social psychology (assimilation, desire to have an "illustrious" name etc.)</i>
5.2.1 Immigrant names borrowed semantically	Surnames borne by "autochthonous" families	1. <i>Semantic association to previously used surnames</i> 2. Assimilation
5.2.2 Immigrant names borrowed phonetically	Surnames borne by "autochthonous" families	1. <i>Phonetic association to previously used surnames</i> 2. Assimilation
5.2.3 Other names	Surname borne by "autochthonous" families,	<i>Assimilation</i>

borrowed by immigrants	often commonly used	
6. Migrated surnames	Surnames used in former countries	Not applicable

The table deserves several comments. First, the numbering of types follows that of categories of etymons given in section 2: the first digit directly corresponds to the number of the corresponding category.

Secondly, the last column presents two different kinds of the generalized answers to *Why*-question:

- objective direct reason that is either clear from the semantics of the etymon (types 1.1, 1.2, 2.1, 2.2 and 3), or association (phonetic or semantic) that the etymon makes to the previously used name (types 1.3, 1.4.1, 1.4.2, 5.2.1 and 5.2.2);
- partly subjective socio-psychological reasons such as assimilation (types 4.2, 5.1.1, 5.1.2, 5.2.1, 5.2.2 and 5.2.3).

It can be observed that types 5.2.1 and 5.2.2 are concerned by both kinds of answers.

Thirdly, the table includes one type of surnames (1.5) that was never mentioned in previous discussion: artificial surnames. This type is commonly present in Ashkenazic Jewish, Russian and Swedish corpuses. In all these cases, one deal with series of surnames constructed after similar patterns and having no relationship to any of the characteristics of their first bearers. As a result, no objective answer to *Why*-question can be suggested on the individual level. In the frame of the scientific approach, one can discuss only the reasons for the creation of whole series and patterns used in them.¹⁷

This classification provides the formal presentation for the principle that the etymology of a surname strongly depends of the ethno-cultural group. Table 2 shows it explicitly for one example: the surname Smith.

Table 2. Typology of the surname Smith in various groups

Group	Type	Explanation
English	1.1	Designates the occupation of the first bearer or his father
Anglo-American	6	Appeared in US after the migration from England without name change
African-American	5.1.1, 5.1.2	Received by the first black bearer from the owner of the plantation
German-American	5.2.1	Translation of Schmidt, the surname used in Germany

The primarily task of the etymological research consists of finding source words and assigning the name to one of the types listed in Table 1. The following sections discuss the methods allowing to estimate whether the suggested etymon and the type of surname are reliable.

4. *How*-question or Differences Between Surnames and Their Etymons

Any etymological analysis claiming to be scientific would be necessarily incomplete if it does not explain *how the surname from its etymon was created*. This question (below referred to as *How*-question), includes an explanation of the phonetic and/or graphic difference that can exist between a surname and its etymon. In other words, all elements present in the surname should be explained.

The answer depends on the category of etymons. For the first three categories cited at the beginning of section 2, it is purely linguistic and it is related to morphology. In numerous European cultures, large number of surnames¹⁸ are drawn by adding supplementary elements before and/or after the source words. These elements can be either actual suffixes (often possessive or diminutive) and prefixes, or words meaning “son,” “grandson,” “descendant of.” They are particularly common in patronymic surnames, cf. northern

English *-son*, southern English, Dutch and German *-s*, Gaelic *Mac/Mc* (son) in Scotland and Ireland, Gaelic *Ó*, Anglicized to *O'* (grandson) in Ireland, Danish, northern German, Norwegian and Frisian *-sen*, Swedish *-sson*, Spanish *-ez*, Portuguese *-es*, general Italian *-i*, Venetian *-ato*, Romanian *-escu*, southern Slavic *-(ov)ič*, Polish *-owicz* and *-ewicz*, Russian and Bulgarian *-ов* [ov], eastern Ukrainian *-энко* [enko], western Ukrainian *-(ч)ук* [(ch)uk]. Specific suffixes are commonly present in surnames drawn from toponyms, cf. German *-er*, Hungarian *-i*, Polish *-ski*, Czech *-ský*, eastern Slavic *-с(ь)кий* [skiy, s'kyy]. Prefixes meaning "of, from" are often present in surnames drawn from general vocabulary words of several cultures: Dutch *van* (also present in names derived from placenames), French *Du-* and *De-*, Italian *De-* and *Di-*. Genetically, Finnish *-nen* was a diminutive suffix with possessive meaning, often expressing association with a place. Later, it represented just an element that could be added to any root. Numerous suffixes, mainly substantive, are present in Slavic surnames derived from common nouns, adjectives and verbs. Apart from very few exceptions, the affixes used when constructing surnames belong to the same language as the etymon. This correlation should be taken into account when evaluating the reliability of suggested etymons.

The creation of surnames by abbreviating patronymic, occupational and other types of expressions are peculiar only to Ashkenazic Jews, mainly from Eastern Europe.

For secondary names (type 4.1), the difference between the old and the new surnames, both used within the same family, are generally not due to affixes. The cases of intentional changes are rare, cf. shortened forms used for illegitimate children of Russian nobles such as *Пнин* [Pnin] from *Репнин* [Repnin] and *Бетской* [Betskoj] from *Трубетской* [Trubetskoy], or the inversion of the order of parts in German Jewish Bachauer from Auerbach. Unintentional changes are much more common. Often the distortion results from an error committed by scribes. Misreading of one or several letters in

the old name yield graphic distortions. The same result is achieved when a scribe simply misspells the old name. The phonetic distortions occur more frequently. In these cases, when being pronounced, the name is misinterpreted by scribes. Sometimes, unconscious will to standardize the surname (that is, replace some of its elements by another, phonetically close, but more commonly present in other surnames) or the effects of folk etymology, also unconscious, can be responsible for these phenomena. When conjecturing that a surname is due to the phonetic distortion, it should be taken into account that the confusion between several sounds depends on the phonemic structure of the vernacular language. This correlation between the answers to *What-* and *How-*questions is important for suggesting reliable etymologies. Several examples illustrate this rule. The devoicing of final voiced consonants is standard to German and Russian and unusual to French and English. As a result, a hypothesis that the final /v/ was changed to /f/ is more plausible in two former languages than in two latter ones. In southern German dialects the confusion between /b/ and /p/, /d/ and /t/, /g/ and /k/ is well known, while it does not exist in numerous other European tongues. The reduction of unstressed vowels is also language dependent. In Belorussian, the morphologically correct /o/ is regularly pronounced (and spelled) as /a/ when unstressed (*akanje*). The same tendency exists in certain Russian dialects, while in other Russian dialects, Ukrainian and Polish such phonetic phenomenon is unknown. If in Spain numerous interchangeable forms with "b" and "v" are found, the confusion between these elements in neighboring France are unusual. Due to the phenomenon of *mazurzenie* in certain Polish regions, a confusion exists between the hissing and hushing consonants.

Migrated surnames (type 6) often undergo deliberate changes. Assimilative tendencies are their main stimulus. As a result, often the final, explicitly foreign, elements are dropped.

This pattern was common in Northern America. Alternatively, these final elements can be replaced by those commonly present in surnames of the ethno-cultural group that represents the local majority. For example, Armenians who lived within the Russian empire outside of Armenia itself, were often dropping their original patronymic suffix *-ian* (spelled *-ян* in Russian) to substitute it with *-ob* [ov], the suffix frequently found at the end of Russian surnames.

Purely phonetic changes—when only the pronunciation of a surname of foreign origin is modified, while its spelling is kept unchanged—are also due to the assimilation. In these cases, all letters are read according to the pronunciation rules of local language. This situation is typical, for instance, to US, England and France for surnames of German and Polish origin that include in their structure the letter “j” or the letter combination “ch” which are pronounced in English and French in a way very different from German and Polish.¹⁹ The change in pronunciation is not necessarily made immediately upon migration: it can also occur in one of the following generations. In these cases, the etymon and the resulting surname are spelled identically, but they are pronounced differently.

For types 4.2 and 5, *How*-question is non-applicable: surnames are borrowed without changes.

5. Other Basic Questions of Etymological Research

Several other additional basic questions are of direct interest for the etymology:

- “*Who*-question”: who chose the surname?
- “*Where*-question”: where did the surname appear for the first time?
- “*When*-question”: when was the surname assigned?

Three possible answers exist for *Who*-question:

- state or local officials (including scribes);
- the first bearer himself or a member of the same family;
- his neighbors.

Even in the absence of the direct sources, in a number of cases some factors allow to suggest a plausible answer based on the information provided by the analysis related to other basic questions. The answer depends on the surname type, that in turn represents the cross-reference of *What* and *Why*-questions. Borrowing of surnames from another group of population, can be either forced by officials (type 5.1.1), or made following the natural will of the first bearer (type 5.1.2). Surnames borrowed internally (type 4.2) are necessarily due to their first bearers. The creation of secondary names (type 4.1) can be due to the family of the first bearer only if it is intentional. Those resulting from changes made unintentionally are due to scribes. Both officials and first bearers (but not neighbors) can be responsible for surnames assigned after laws prescribing surname adoption (artificial surnames; type 1.5), following migrations (types 1.4, 5.2 and 6), calques (type 1.3). Only surnames that arose in the most natural way, that is, after the transformation of nicknames into hereditary family names, are ultimately due to neighbors. They cover types 1.1 (occupational), 1.2 (derived from personal characteristics), 2 (patronymic and metronymic) and 3 (drawn from placenames). The same types can be also due to officials or the first bearers. It is also clear that all contemptuous surnames are not due to their first bearers. On the other hand, prestigious names were most likely chosen by the bearers themselves. The simultaneous presence in the same place of a series of surnames with particular suffixes (*How*-question) can sometimes testify about the work of officials. Surnames assigned by officials are mainly based on the official language of the country. The existence of a series of surnames drawn from a language unusual to the area can be due only to the imagination of officials. On the other hand, those taken by people themselves or nicknames invented by neighbors are generally taken from the vernacular dialect.

A meticulous analysis of the chronology and geography of occurrences of a name can make it possible to

find answers to *Where* and *When*-questions, respectively.²⁰ For names that are more than two centuries old, historical documents often do not allow to discover without ambiguity the area of the creation. It can appear that the same name is used in two or more areas. If one considers that all bearers of this name are related (monogenetic name),²¹ then the area where its occurrence is the earliest is likely to be its place of origin. This criterion is not precise. Due to the paucity of historical references, the lapse of time between the creation of an appellation in a region and its first mention in available sources can be important. If, during that period, the name was brought to a new region, there is a chance that its first documented occurrence in this new area could be older than that in the area from which it originated. Another criterion comes from demographic statistics. If a name is used in several areas during the same period, it is more likely to originate from that area of the most frequent use.

The more reliable are the answers to *Where* and *When*-questions, the more plausible would be the knowledge of whether the name in question is monogenetic or polygenetic. The first term applies to appellations that appeared only in one particular place at one specific moment for one particular family, that is, everyone with this surname has the same single progenitor. Polygenetic surnames are those adopted by unrelated families. For them, the same unusual semantics (*What*) is not plausible. Moreover, independent branches of these names can, in principle, derive their surnames from distinct source words. On the other hand, for monogenetic surnames, an etymology which seems strange might, indeed, be correct. Similarly to *What*-question, in answering *How*-question, there is an important methodological difference between monogenetic and polygenetic appellations. For example, the same unusual phonetic distortion is not plausible in a name shared by unrelated families. On the other hand, for monogenetic surnames, a distortion which seems strange might, indeed, be correct.

The answer to *Where*-question is very important in establishing reliable suggested etymon, that is, the answer to *What*-question. Its significance is particularly evident in surnames derived from placenames: The toponym suggested as a source for a name should be known in the area in question; that place either should be situated close enough to the locality where the surname appeared for the first time or it should correspond to an important settlement, such as a large city.

Some names sound alike. If both occur in the same area, this may suggest that the families bearing these names descend from a common ancestor. If the source of one of these names is clear, inferences may be drawn to establish the origin of the other, secondary, surname that arose from the first one after a distortion.

The answer to *When*-question can also be crucial for evaluating the reliability of the suggested etymon. For the source of the surname to be correct, it should not be an anachronism. The surname cannot have been based on new words (or on new meanings of old words) which appeared after the time of the surname adoption.

6. Genealogical Sources

The methodology of natural sciences - consisting in corroboration or refutation of advanced, never-definitive, theories with empirical data - can partly be applied to etymological research as well. Historical records that date from the era of name creation are the best sources for establishing reliable etymologies. They can be considered the empirical data for the science of onomastics. Contrary to considerations presented in the previous section that corresponded to the history of names rather than that of families, these sources provide information concerning specific families. Records can include the given names of various members of the family, occupations, names of native places and places of residence. This information can either

corroborate or refute the answers to *What-* and *Why-*questions for three important categories of surnames, namely those derived from given names, occupations and toponyms. It is of paramount importance if the literal meaning is ambiguous, and does not point to the etymology directly. Table 3 presents several examples taken from Polish Jewish civil records. All of them date from 1820s-1830s, that is, the period that followed immediately the law of 1821 forcing adoption of surnames by Polish Jews. In these cases, the information present in civil records directly allows to establish the etymology.

Table 3. Examples from Polish Jewish civil records

Surname	Literally meaning	Civil record information
Felman	fur / skin + man (Yiddish)	occupation: tanner
Hepner	hops + suffix <i>-er</i> with root vowel alternation /o/ > /e/ (Yiddish)	occupation: dealer in hops
Krugman	tavern/jug + man (German)	occupation: tavern keeper
Krupnik	grain + suffix <i>-nik</i> ; mead; barley soup (Polish)	occupation: dealer in grains
Obarzanek	bagel (Polish)	occupation: baker
Rybacki	fisherman's; fishmonger's (Polish)	occupation: fishmonger
Zyngier	singer (Yiddish)	occupation: cantor in a synagogue
Herszberg	deer (or given name Hersz) + mountain (German)	first name: Hersz
Prynczstein	prince + stone (Yiddish, German)	mother's first name: Szprynca
Kalina	guelder rose (Polish)	native village: Kalinowo
Korobczyk	box + suffix <i>-czyk</i> (Polish)	native village: Korobiec

Direct testimonies about the origins of names are the most precious sources in etymological research. Unfortunately, generally they can be found only for surnames adopted during the last centuries. In this case, neutral

documents - such as name adoption lists that recorded the event without explaining the reasons of choices of specific surnames - are totally reliable. On the other hand, explanations provided by those who made the choice are to be read with caution: these people could sometimes have psychological reasons to modify the reality.²²

Indirect oral testimonies transmitted via family stories deserve much more caution. This source of information is not psychologically neutral *par excellence*: people have emotional attitude to their proper roots and as a result they have a tendency to "ameliorate" the etymology of their surnames. The amelioration is often manifested unconsciously by fitting the etymology to the aspirations that people have about their roots. Evidently, derogatory interpretations are seen as objectionable, hence incorrect. The rather reluctant acceptance of derivation of surnames from the names of small villages is also common: such etymologies may be seen as too prosaic. People may want to find an unusual story concerning their families and will easily accept interpretations that flatter the imagination. These unusual stories often become attached to the surname adoption: the surnames appear to be the oldest elements preserved in the families due to the paucity of available documentation. As a consequence, people can easily accept an exotic etymology that distinguishes their family from others. If during the last hundred years this tendency has been applicable to numerous representatives of middle class, for aristocratic families it has been many centuries old. The so-called *Velvet Book* of Russian nobility, compiled in 1685 and encompassing family stories that were much older, can serve as an illustration. The general aspiration of that time was to have an ancestor who came to Russia from abroad. As a result, many families, even those bearing surnames the meaning of which was self-evident for any Russian, were suggesting foreign etymons relating them to West Europe, Scandinavia, Greece or even ancient Rome. Only in a few cases, mainly dealing with recent migrations, *Velvet Book* provides reliable information.

Among the examples are two families with authentic Scottish roots: *Лермонтов* [Lermontov] from Learmonth and *Хомутов* [Khomutov], a Russified form of Hamilton.

An important gap in generations often existed between the first bearers of surnames and those who faced for the first time the problem of determining their etymology. In these cases, interpretations often represent the fruits of the imagination. When individuals asked their parents or other relatives about their family origins and received no definitive answer, many looked in dictionaries or on maps. Numerous folk etymologies of names have entered into family stories in this way. One psychological phenomenon explains the persistence of these folk etymologies in later generations. When a small child, or even a teenager, questions his parents, his response to the received information may differ substantially from the importance attached to it by the parents. The latter might pay little attention to it; the entire story might be just the first thought that came to mind in the moment of questioning, and it might easily be forgotten later. On the other hand, the child might assume this information to be the only possible interpretation because it was spoken by the ultimate authorities, the parent. What may well have been a hypothesis has now taken on the character of a fact.

Several general tendencies of folk etymologies can be discerned. Firstly, one of the standard techniques is to look directly at the name root. Such etymologies can be created by different persons as they independently consult the same source, such as dictionary. Using this method, without strong genealogical support, one can ignore the facts of borrowing of surnames and other nuances presented in Table 2. Paying no attention to the suffix present in the name, one will be often unable to distinguish surnames drawn from the general lexicon from those derived from placenames (that are in turn derived from these general words). Secondly, hypothetical etymologies may be influenced by their final or initial elements. For example, knowing that numerous surnames came from Brittany

end in *-ec* or *-ic*, one can erroneously suggest similar origin for a number of appellations that actually arose in other French regions such as Amalric.

7. General historical background

The consideration of a number of general features concerning the historical background of the surname adoption is of paramount importance for establishing a reliable etymology. The data can be obtained from the analysis of historical documents dealing with the time and the area of name creation, that is, once the answers to *Where-* and *When-* questions are known.

First of all, this includes the detailed knowledge of the legal aspects of the process of acquiring surnames. The existence of laws making surname adoption mandatory provides the answer to the question why was a name to be chosen. Such circumstances are often responsible for specific features of the corpus of names: the use of the official language, participation of state clerks in the assignment, creation of standard patterns allowing to construct large series simultaneously. If surnames appeared after the putting in practice of laws, the knowledge of official naming traditions valid in the same ethno-cultural group just before the surnaming process is important: the newly adopted surnames can inherit one part of these older customs. On the other hand, if during the period of the name creation no law existed making surnames mandatory, officials could not be responsible for this name adoption.

Secondly, in many cases historical data can be used to validate that the suggested etymon is in principle appropriate. The hypothesis of a toponymic derivation is refuted if it appears that the locality in question was nonexistent at the time of the creation of the name. It is strengthened if the representatives of the ethno-cultural group under considerations had lived there. By the same token, for a given name to be plausible as the source of a surname, it should

have been used during the same time period. If a word that refers to an occupation presumed to have been the basis of a surname, it is important to demonstrate that people of that ethno-cultural group were involved in the occupation at that time. The language to which belongs the suggested etymon should be either used officially, spoken vernacularly at the moment of the name creation, or the language of culture as, for example, Latin and Greek for Christians. If the surname differs from its etymon by the presence of specific suffix or prefix, the data of historical philology should be used to show that this morphologic element was used in that area at the time when the name originated.

A reliable etymology implying that the surname was borrowed should prove that the surname in question was indeed borne either by representatives of the same ethno-cultural group or members of politico-economical dominant group.

For names that are due to immigrations, their provenance from a specific foreign country would be strengthened by the general information about the existence of migrants from the country in question.

8. Statistical analysis

8.1 Types of statistics

For the etymological research concerning one particular surname, the information obtained from the statistical analysis of other surnames borne during the same period by members of the same ethno-cultural group can be of great help. In principle, theories concerning different names are independent. Still one can formulate general conclusions which - on a statistical basic - can be valid for various individual cases (though of course, a number of names will be totally out of any generalizing theory). These conclusions are of particular importance for names the adoption of which was forced by law since in these cases the same pattern could be used to construct series of names. It is crucial for artificial

surnames (type 1.5). For them, as it was already discussed in section 3.6, *Why*-question can never be answered and as a consequence, without resorting to the statistics on other surnames adopted at the same period, any hypothesis concerning the etymology of one particular surname remains speculative. The statistical results are least valid for names that appeared via a gradual transformation of individual nicknames into hereditary appellations.

The following calculations can be performed:

A. Statistics of affixes. This calculation of the relative frequency of presence of various suffixes, prefixes and infixes in surnames can be used to corroborate the answer to *How*-question. The more this morphologic element at the beginning of the name under analysis (prefix), at its ending (suffix) or inside it (infix) is common in other surnames belonging to the same population group, the more the hypothetical answer to *How*-question is plausible. A high frequency of use of one particular affix in a small compact area can be generally due to two different situations. Firstly, it can correspond to standard official naming patterns of that region during the time when no surnames existed. This is the case, for example, of Russian possessive *-OB* [ov], *-EB* [ev] and *-IH* [in]. Depending on the final sound of the father's first name, one of these three elements was added to that first name to construct a patronymic, a name category regularly present in official Russian documents during the period when no hereditary appellations were used yet. Secondly, the high proportion of surnames with the same peculiar affix can reveal the fact that the whole series was created by the same clerk. This is particularly plausible if this morphologic element is not common elsewhere. Consequently, this data sheds light on the *Who*-question.

B. Statistics of languages of the etymons. This calculation of the proportions of etymons taken from various tongues can be used to corroborate the answer to *What*-question. If two concurrent hypotheses explain the origin of

the same surname from etymons belonging to two distinct languages, the language responsible for the creation of a larger number of other surnames is in more favorable position for the surname under consideration.²³ As it was already discussed in section 5, a correlation exists between the language of the etymon and the answer to *Who*-question. Consequently, this statistical data can also shed light onto that question.

C. Statistics of semantic categories. This calculation of the proportions of etymons belonging to various semantic groups (occupational terms, words designating various human qualities, male or female given names, placenames, names of animals, plants or minerals, abstract terms etc.) can be used to corroborate the type of a surname, that is, the cross-reference of the generalized answers to *What*- and *Why*-questions. The larger the proportion of appellations of the suggested type, the more plausible is the conjecture. Similarly to two previous statistics, the presence of a populous series of surnames belonging to one specific semantic category, unusual for other places, can reveal the fact of a simultaneous creation of the whole series as a result of the activity of the same official (*Who*).

D. Statistics of phonetic changes in secondary surnames (type 4.1). These phonetic phenomena are not specific to surnames being relevant for the language as a whole. As a result, they are mainly known from general linguistic studies.²⁴ The information provided by these sources is not, nevertheless, exhaustive: the analysis of secondary surnames can sometimes shed light on general dialectal features that cannot be discerned using other linguistic tools. In the specific context of etymology of surnames, statistics on phonetic changes allows to estimate the reliability of the answer to *How*-question. Indeed, a suggestion that a name results from another after a particular sound confusion, becomes more plausible if the same distortion is found in numerous other surnames of the same ethno-cultural group. On the other hand, if the surname under consideration

represents the unique example of the phonetic change in question, the hypothesis about its etymology becomes implausible and can be called *ad hoc*.

E. Statistics of names retained or chosen by immigrants
This calculation proceeds in several steps:

- Compile a representative list of surnames, used before and after the migration, by people from the same ethno-cultural group who migrated during the same period to one specific country;
- Compare the new and the old names allocating them to one of the types listed in Table 1 that deal with immigrants (1.4.1, 1.4.2; 5.2.1-5.2.3; 6) and calculate this way the distribution by type;
- For surnames of type 6, that is, those whose etymons are surnames used in the former country, calculate the frequency of use of various patterns such as:
 - neither pronunciation nor spelling change;
 - only spelling changes (pronunciation kept);
 - only pronunciation changes (spelling kept);
 - dropping of suffixes and prefixes;
 - addition of suffixes and prefixes.

This statistical analysis makes it possible to discover general trends peculiar to the group of immigrants under analysis. As a result, this information can be useful for social psychology. For onomastics itself, it provides data allowing to estimate the reliability of suggested etymological hypotheses. If the name used by an immigrant family in the former country is unknown, the information provided by the statistical analysis of other surnames from the same group (and more precisely, the most common types and patterns) shows in what direction the researcher should go: for example, looking for phonetic or semantic association that the newly acquiring name was making to that borne before the migration or seeking for specific foreign suffixes dropped in the new country.

8.2 General principles

To obtain reliable results, the corpus of surnames taken for statistical calculations should possess several properties. First, it should include only authentic surnames. This principle sounds evident, still, for example, when studying Jewish Ashkenazic names, I was able to observe that certain researchers invent peculiar theories for appellations that never existed but appeared in printed sources as a result of typographic errors or misreading of handwritten documents.²⁵ To deal with names taken from historical sources, it is preferable to corroborate their existence by several independent references. Secondly, when making statistics on affixes, languages and semantic categories every family should be taken into account only once. As a result, these calculations should exclude secondary surnames if their etymons (primary surnames) are already taken into consideration. Thirdly, surnames for which the characteristics under analysis (affix, language, semantic group, phonetic change) is doubtful should be ignored. Indeed, the main reason for these calculations is to provide reliable statistical information useful to corroborate hypotheses concerning questionable etymologies. Consequently, for surnames retained for the calculations, the characteristics on which the statistics is made, should be relatively clear. Formally speaking, this principle is a real problem since, as it was stressed out at the beginning of section 1, for any surname its etymology can never be definitively proved: we do not know its elements (affixes, language etc.) for sure, they continue to be hypothetical. As a result, the whole method sounds rather dubious: when using it, one tries to corroborate one hypothesis by statistical analysis of other hypotheses. In practice this formal objection is, nevertheless, significantly relieved: fortunately, the etymology of numerous surnames is generally self-evident. Fourthly, one should not forget the principle standard to any statistics: the larger is the sample, the more are the chances to reveal some objective trends rather than to fall on contingent results.

The possibility of corroborating a hypothesis concerning one specific name by using statistical analysis of other names relies upon one underlying postulate that is general to humanities: even if every individual is specific, he shares numerous elements of his behavior with members of the same ethno-cultural group to which he belongs. In the context of onomastics, this means that when creating a surname, various persons from the same group can be guided by similar ideas and, as a result, surnames of different families can be created using identical patterns. The more the group in question is homogeneous, the more the results of the statistical analysis are reliable. Several characteristics allow to estimate the level of its onomastic homogeneity. Firstly, all members of this group should live on a continuous territory. The smaller is the region, the more homogeneous is the group. Secondly, during the surname adoption they should speak the same vernacular language(s) and have the same cultural language (if any), used in specific contexts. Thirdly, for all of them, the legal conditions of the adoption of surnames should be identical, and, if the acquirement of the hereditary appellations was forced by law, the official language of the country should be the same. These three conditions are mandatory to allow a reliable statistical analysis.²⁶ Only if all of them are satisfied we can really speak about the same *ethno-cultural* group, the term that was many times used throughout this paper. Fourthly, the shorter is the period during which they acquired their surnames, the more homogeneous is the group. The belonging to the same social strata represents the fifth important criterion concerning the homogeneity. Note that the above criteria are not really independent, and, for example, it is a common situation when the nobility, that is, the highest social strata (criterion 5), acquires surnames several centuries before other social groups (criterion 4) for which a special legislation is being applied (criterion 3); some of these appellations can be drawn from a language spoken only by the nobles (criterion 2).

A judgment about the homogeneity is better expressed in relative terms. For example, it is clear that modern Serbs are more homogenous than modern French since the latter encompass numerous persons whose family names originated outside of the general context of the French language (western Brittany, Italy, Portugal, Spain, Northern and Central Africa, etc.), and, therefore, do not conform to the second and the third conditions formulated above. During the same period, one part is obviously more homogenous than the whole. For instance, the onomastic homogeneity diminishes in the following chain: Jews from the Lublin area—Jews from southern Poland—Polish Jews—Ashkenazic Jews—Jews. For the last element of this chain, any statistical analysis makes no sense since all the principal criteria formulated above are not applicable to it. For the second to last element, the analysis of this kind is also problematic since – even if the time of the adoption, social and legal conditions and even the vernacular language (Yiddish) are often similar for various groups of Ashkenazic Jews who lived on a continuous area of Eastern and Central Europe—the official languages of various countries were different (principally Russian, Polish and German), this factor creating important onomastic differences. Similar observations can be made concerning the following chain: American Sicilians—American Italians—Americans.

A general rule can also be formulated for the same territory in different periods. Normally, the older are the times taken into consideration, the more homogeneous is the corpus of surnames. This is particularly true for the period that follows immediately the time of the adoption of surnames.²⁷ For recent times, migrations contribute in a natural way to make the corpus more heterogeneous. Generally, they are more influential than another natural phenomenon tending to the opposite direction, that of the homogenization: disappearing of rare surnames in families that have no male descendants. This rule has, however, several important exceptions. For example, the modern corpus of Polish

surnames is much more homogenous than it was before the Second World War and encompassed appellations of numerous people belonging to German, Ukrainian and Jewish communities of Poland.²⁸

Conclusion

The etymology of surnames is a separate interdisciplinary domain that uses linguistic, historical and genealogical data and methods, as well as its own methods. In the present paper, an attempt is made to provide an approach that could be called scientific. It leans on several ideas. Firstly, the main aim of the etymological research is the answer to the question about what is the source word (*What*-question) that allows to suggest an etymon. In principle, the definitive objective knowledge is possible since for any appellation its etymon does certainly exist. Secondly, any suggested etymon remains a hypothesis that can rarely be definitively proved. In other words, the proposed solution can be, in principle, 100 per cent objectively true, but we cannot be 100 per cent sure that it is indeed the case. In this situation, a scientific approach asks to provide a conjecture that would be the most corroborated by various factors among which are:

- Answers, correlated with the etymon and between them, are provided to other main questions of the etymological research concerning the surname under analysis: *when* and *where* the assignment took place, *who* chose the appellation, *how* the surname was created from the source word, *why* the etymon was chosen to construct the surname. For all these questions (except for the last one), the answer can, in principle, be 100 per cent true. For *Why*-question, understood in its narrow sense asking for the direct reasons only, there are situations for which the answer can also be totally objective (occupations, personal characteristics, given names used in the same family, places of residence or origin, and the existence of associations to names used previously);
- Results of the statistical analysis of surnames of other families belonging to the same ethno-cultural group can

corroborate the etymon and the answers to *Why-*, *How-* and *Who-*questions;

- Knowledge concerning the general historical and linguistic background of the time of the surname adoption that can corroborate the etymon and the answers to *Why-*, *How-* and *Who-*questions;

- Finding of the family genealogical data concerning the time of the surname adoption (such as proper names, occupations, places) that are correlated with the suggested etymon;

- Existence of the testimony concerning the surname adoption.

The presence or the absence of these elements, allows to establish the hierarchy of reliabilities of various hypothesis concerning etymologies that can be used to compare either concurrent theories concerning the same name or conjectures dealing with different surnames. If all these factors are present simultaneously then the etymology is the most reliable. If none of them is provided then the etymology is purely speculative (though, in principle, it can be accurate). Other situations are intermediate. On an individual level, the importance of the factors in the list above increases. For example, the correlation provided by two last factors contributes to the verisimilitude of the hypothesis more than all the other factors taken together. For certain other cases, the increase of the importance is a direct consequence of the fact that one factor implicitly suggests the presence of another. For example, without the answer to *Where* and *When*-questions (the first factor in the list above), neither the historical background can be established, nor the genealogical data concerning the family during the period of surname adoption can be found. Similarly, the statistical analysis and the knowledge of the general background are of real interest only if they corroborate not only the fact that the language of the etymon is appropriate, but also the answers to *Why-*, *How-* and *Who-*questions (that are, therefore, presupposed to be provided too).

The correlation found on any level is not necessarily total and can be of different degree. For the most important level, only the testimony that is impartial and direct is really reliable.²⁹ For the following level, the correlation between the etymon and the available genealogical data can also be direct or indirect, the former being much more reliable than the latter.³⁰ The results of the statistical analysis correspond to various degrees *par excellence*. For the most basic level, the answers to some questions can be partial (for example, only a very general area or the time period suggested for the surname adoption); to other questions they cannot be provided at all. From one side, these observations show the difficulties of a formal approach in comparison of two hypotheses (for example, no weight of various factors and no exact mathematical degree of absolute correlation can be objectively introduced). On the other side, they also illustrate how, in principle, this comparison can proceed: by comparing relative correlations, by giving preference to direct and impartial sources and to conjectures that are better corroborated by the statistical analysis. It is also clear that the smaller is the number of independent unusual conjectures contained in the global etymological hypothesis, the larger is the verisimilitude of this hypothesis. For example, a theory that explains a surname as (1) drawn from a common word taken from a language unusual for the region where the surname was adopted, to which (2) a suffix was added that is also uncommon to that region, has very little chances to be true.

The basis for the objective knowledge in etymology of surnames is provided by the exact answers to *Where-* and *When-*question, the direct and impartial genealogical information (but not by family stories), and data concerning the general historical and linguistic background of the area and the time of the surname adoption.

Once suggested, a hypothesis concerning the etymology of one particular surname can be later corroborated or refuted by

the discovery of additional information dealing with one of the factors listed above: genealogical data, new knowledge concerning the historical or linguistic background, compilation of a statistically representative list of surnames belong to the same ethno-cultural group, more detailed answers provided to the main questions of the etymological research. For example, if a surname is supposed to be derived from a toponym and only the general region of its origin is known, a more narrow localization of the area where it arose would be of crucial importance to evaluate the verisimilitude of the conjectured etymology.

Notes

¹ Beider, A. *A Dictionary of Ashkenazic Given Names: Their Origins, Structure, Pronunciation and Migrations*. Bergenfield, NJ, Avotaynu Inc., 2001.

² The same is true for placenames.

³ The meaning of the expression "ethno-cultural group" is discussed in section 8.2.

⁴ In this paper, the word "he" should always be understood as "he/she." Similarly, "his" means "his/her."

⁵ For example, despite important linguistic, philological and historical traditions of the corresponding areas, the first scholarly study of French names was published only in 1951, those of Czech, Russian, Ukrainian and Belarussian surnames were compiled only during the 1960-70s. For Poles and Jews from Eastern Europe, the first books from that category were written during the 1990s. The first comprehensive reference work on Spanish surnames appeared in 2001.

⁶ Here, I totally support Karl R. Popper's ideas concerning three worlds: (1) objective physical world; (2) subjective thoughts and experiences; (3) products of human thoughts (cf., for example, *Objective Knowledge*, Oxford University Press, 1979). Popper stressed out that the study of the elements of World 3 can shed light onto our knowledge of World 2 and strongly criticized the opposite approach. His arguments were directed against the philosophers who – as Dilthey and Collingwood – considered that the central task of a study in humanities consisted of a kind of intuitive identification with other individuals allowing to relive the experiences of others.

⁷ The concepts presented below are not exhaustive from the geographic point of view. They will be limited to the analysis of European and American cultures; peculiarities of Asian and African surnames are beyond the scope of this paper.

⁸ Cf. the testimony made by an Austrian official who explained that he assigned the surname Weinstein to a Jew who cried (German verb *weinen*)

and groaned (*stöhnen*) during the surnaming procedure (Franzos, K.E. "Namensstudien." *Halb-Asien* 5:1:125–149. Stuttgart, 1888). We know, however, that (1) the same name was assigned to numerous other independent Jewish families in the Hapsburg Empire, (2) a large number of other surnames starting in *Wein-* (wine) or ending in *-stein* (stone) were invented at the same time by various Austrian clerks. As a result, the odds that the author of the testimony in question simply played with words in his description are very high.

⁹ Many surnames could be, in principle, attributed to the fifth category since, for example, Black, Field and Miller are common English surnames too. For the specific needs of the etymological research, the choice between the first and the fifth categories for surnames of this kind is, however, of no importance.

¹⁰ One family of Italian origin can be used as a curious example. In Italy, the name was Puni and when a branch of this family migrated to Russia, it was transformed to Пунн, the Cyrillic form that corresponds to both the direct transliteration and the phonetic transcription of the original Italian name. When at the beginning of the 20th century a member of the same family, a Russian artist, migrated from Russia to France, his name became Pougny, that sounds in French similarly, let alone the shift of the stress from the first to the last syllable.

¹¹ This principle represents the basis for the application of statistical analysis (see section 8.2).

¹² The described situation resembles that of historiography. Indeed, rare are the historians (for example, Langlois, Seignobos and other representatives of the so-called *Methodic School* that existed in France about one century ago) who intentionally restricted their intellectual activity to factual information. The search for answers to the question *Why?*, that is, attempts to explain the reasons behind different events, are much more common.

¹³ See section 8.

¹⁴ The possibility that a surname reflects the occupation of the first bearer's father is more plausible and it is doubtless when the surname ends in a possessive suffix, as in the Italian and Russian examples given above.

¹⁵ See section 8.1, statistics E.

¹⁶ The answer would explain the individual choice made between the first, the fifth and the sixth categories.

¹⁷ See additional details in section 8.

¹⁸ Properly speaking, this statement is very often valid not for surnames, but nicknames that later gave rise to surname; cf. section 1.

¹⁹ Other examples, such as Zimmermann and Sommer, were considered already in section 2.2.

²⁰ Properly speaking, in most European cultures surnames evolved over a long period of time during which gradual transformation of personal nicknames (bynames) into hereditary fixed appellations took place. As a result, it seems that the answers to the question *where* the surname first appeared and *when* it was adopted are generally impossible to determine. In our approach, however, as it was already said in its beginning, this formal difficulty does not exist: when speaking about surnames drawn from

nicknames, different questions are actually addressed not to surnames, but to nicknames from which they are derived. For every nickname, the place where and the time when it was first given are unambiguous concepts (even if they often difficult to determine in practice).

²¹ See the next paragraph.

²² Cf. section 1.

²³ In this statement it is implicitly taken for granted that according to other criteria both hypotheses are equally plausible.

²⁴ For example, this is true for the cases cited in section 4 (final devoicing, reduction of unstressed vowels, confusion between hissing and hushing consonants etc.).

²⁵ The validity of this observation is based on the analysis of other historical documents in which the same individual or family are mentioned.

²⁶ If the principle of common territory is indeed obligatory, in some specific cases the statistical analysis can also make sense even if one of the two other conditions are not satisfied. For example, if a family established for several generations in a small American town is not aware about the country of origin of its ancestors and its surname is too ambiguous to provide this information, the statistical analysis of surnames used in other families of the same town can be helpful. Note that in this situation the second condition (same language at the moment of surname adoption) is not satisfied.

²⁷ Evidently, this rule cannot be directly applied to cultures for which the adoption of surnames was gradual and lasted several centuries.

²⁸ Of course, properly speaking the pre-war Poles could not be considered a group for which any statistical analysis of surnames was appropriate as several necessary conditions were not satisfied (common legislation, identical language during the period of surnaming etc.).

²⁹ See section 6.

³⁰ See the discussion of metonymic occupational surnames in section 3.3.