Preserving Nuance in Chemical Nomenclature Translation

By Roger Sayle

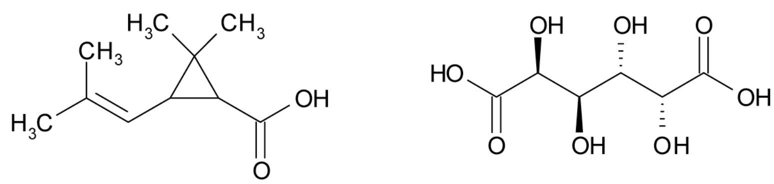
One of the great achievements in the field of chemistry has been the development of its nomenclature. That the world around us contains a multitude of substances with differing properties has been known for centuries, but it is only with the relatively recent arrival of molecular chemistry that scientists are now fully able to classify and characterize different compounds. Where once medieval alchemists could distinguish perhaps a thousand different substances, modern analytical chemistry can differentiate an unlimited number.

The need to communicate the structures of these molecules has led the field to develop its own specialized terminology and complex grammatical rules that attempt to give each molecule a unique name. As the science of chemistry continues to advance, so does its nomenclature. Hence, one of the many challenges faced by translators is the dynamic nature of chemical terminology. For instance, many names that were used 20 years ago differ from those used 10 years ago, which differ again from those used today. Likewise, the names used to identify chemicals can also vary among scientific disciplines. Scientific articles describing the same drug may use one name in a medical journal, a different name in a pharmacology journal, and yet another in an organic chemistry journal. An understanding of the origins of these differences can often help when translating a document or interpreting information from one language to another, where frequently multiple valid terms exist for a single molecule in both the source and target languages.

Types of Names

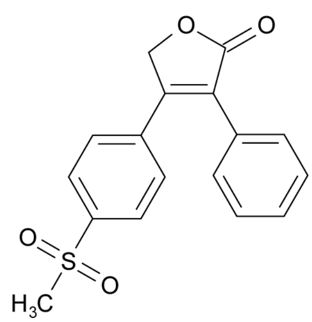
To help readers start to understand why there is so much variation when

Figure 1



Chrysanthemic Acid (菊酸)   Mucic Acid: (粘液酸)

Figure 2



- English: 3-(4-methylsulfonylphenyl)-4-phenyl-2H-furan-5-one
- Dutch: 3-(4-methylsulfonylphenyl)-4-phenyl-2H-furan-5-on
- English: 3-(4-metilsulfonilfenil)-4-fenil-2H-furan-5-ona
- French: 3-(4-méthylsulfonylphényl)-4-phényl-2H-furan-5-one
- Kana: 3-(4-metylsulfonylfenyl)-4-fenyl-2H-furan-5-on
- Polish: 3-(4-metylosulfonylofenylo)-4-fenylo-2H-furano-5-on
- Hungarian: 3-(4-metilszulfonilfenil)-4-fenil-2H-furán-5-on
- Russian: 3-(4-метилсульфонилфенил)-4-фенил-2H-фуранон-5
- Japanese: 3-(4-メチルスルホニルフェニル)-4-フェニル-2H- フラン -5- オン
- Chinese: 3-(4- 甲基磺酰苯基)-4-苯基-2H-呋喃-5- 酮

it comes to naming chemical compounds, let me start by introducing a broad categorization of the major naming styles used in chemistry. This list is not exhaustive, as many small branches of chemistry continue to use their own terminology internally. However, the descriptions will highlight the tensions between naming conventions and explain why the field continues to evolve.

Common Names

Common names are those of ancient origin that predate any attempt at standardization. They include words like water, gold, copper, and urea. For example, through basic necessity, all human cultures had a need to denote water, and the word used to represent this element is universally among the oldest words in any language. It is interesting to observe that English and Dutch use the same word for water, reflecting their shared Anglo-Saxon roots. Likewise, Chinese and Japanese use the same character for water, 水, which predates the divergence of the two languages some 3,000 years ago. Similarly, many Latin-based languages share similar terms for water: *aqua* (Latin), *agua* (Spanish), *agua* (Portuguese), *acqua* (Italian), and *apa* (Romanian).

A common pitfall for translators is when names are “common” in one language but not in another. An example is Japanese, where most chemical nomenclature is translated phonetically using katakana—the more angular, less commonly used of the two Japanese syllabaries—but common names have their own Kanji characters. An English-Japanese translator needs to be aware that individual compounds, like chrysanthemic acid and mucic acid, are exceptions to this rule. (See Figure 1). Due to the lack of regularity, common names must be translated via a dictionary.

Systematic Names

Systematic names use internationally standardized terminology and grammar to constructively describe molecules. Examples include methane, ethane, propane, and butane. These “rational” names lie at the opposite end of the chemistry nomenclature spectrum from common names. For example, in fully systematic nomenclature, “oxidane” is used instead of “water.”

A truly remarkable achieve- ➡

ment of the International Union of Pure and Applied Chemistry's (IUPAC) international standardization of fully systematic names is that it uses what is in effect a phonetic language. Although compound names may look very different when written down, for example, using non-Latin character sets, the spoken or pronounced form is often recognizable to chemists around the world. A Japanese speaker at an academic conference reciting an IUPAC systematic name will often be understood by a native English, German, or Russian speaking chemist. (See Figure 2 on page 23.) A person does not have to be a linguist to recognize that the combinations above all name the same chemical structure.

Trade names pose an interesting challenge for translators and interpreters, as the same medicines may be sold under different brand names geographically, even within the same language.

Traditional Names

Traditional names are the result of early attempts to standardize chemistry in the 1800s. The birth of modern chemical nomenclature is considered to be the pioneering work of Antoine Lavoisier and his colleagues, who first published a list of chemical elements and a set of naming recommendations. Lavoisier's work introduced a number of underlying principles that were shared but translated differently between languages.

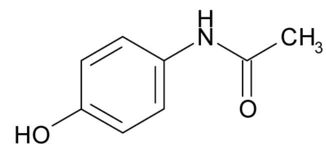
For example, look at the terms formic acid and lactic acid. The English "formic acid" is derived from the Latin word for ant, *formica*. The French, *acide formique,* Spanish *ácido fórmico,* and Romanian *acid formic* all follow the same Latin root. In German it is called *ameisensäure* (the German for ant is *ameisen*); in Polish it is *kwas mrówkowy* (from the Polish word for ant, *mrówka*); in Hungarian it is "hangyasav" (from the Hungarian word for ant, "hangyák"); and in Swedish it is *myrsyra* (from the Swedish word for ant, *myra*). The name "lactic acid" in English is derived from milk, leading to the German *milchsäure* and the Swedish *mjölksyra*. The English name "butyric acid" is derived from "butter," leading to the German *buttersäure* and the Swedish *smörsyra*.

Modern Names

Modern names, such as those given by IUPAC and the Chemical Abstract Service (a division of the American Chemical Society), are perhaps the most common names found in scientific literature, patents, supplier catalogues, and government legislation. These are effectively a pragmatic compromise between the unambiguous ideal of "systematic names" and the national "common names" and "traditional names" historically used by chemists. As with adopting the metric system, the difficult task of reforming and standardizing chemical nomenclature cannot be done overnight. Indeed, the ongoing process is gradual and has already taken several decades. IUPAC's rules for organic chemistry are regularly revised and published, most recently in 1957, 1965, 1971, 1979, and 1993 (and a new manuscript is in preparation). Chemical nomenclature is becoming more systematic with each iteration.

For example, when I was in high school in the U.K., I was taught to use the name "propan-2-one" instead of the former traditional name, "acetone," which was used by the previous generation. I was also taught to spell "sulphur," whereas the current generation is now being taught to use the word "sulfur." IUPAC allows both the traditional prefixes "chloro" and "bromo" and the equivalent system-

Figure 3



- **IUPAC Name:** N-(4-hydroxyphenyl) acetamide
- **Traditional name:** N-(p-hydroxyphenyl) acetamide
- **Systematic Name:** N-(4-hydroxyphenyl) ethanamide
- **Africa, Asia, Central America, and Australasia Trade Name:** Panadol
- **U.S. Trade Name:** Tylenol
- **USAN:** acetaminophen
- **INN:** paracetamol

atic prefixes "chloranyl" and "bromanyl." (IUPAC currently continues to prefer the former, but the switch to the systematic form is probably inevitable within a generation or two.)

Trade Names

Trade names are the proprietary marketing names given to chemicals by manufacturers and vendors. Examples of trade names include Zantac, Lipitor, and Viagra. These names pose an interesting challenge for translators and interpreters, as the same medicines may be sold under different brand names geographically, even within the same language. The headache tablet sold as Tylenol in the U.S. is sold under the name Panadol in much of the world, though it is commonly referred to by its scientific name, paracetamol, in the U.K.

Scientific Names

Scientific names are nonproprietary, and are given to drugs and other chemicals by standards organizations as a convenient shorter name for a widely used (i.e., marketed) compound. These names are commonly used in scientific publications and prescriptions to avoid specifying a particular trade name or formulation. Examples of scientific names include ranitidine, atorvastatin, and sildenafil (Zantac, Lipitor, and Viagra, respectively). Unfortunately, although efforts are made to standardize drug names internationally, the names are assigned by different authorities, so they sometimes differ. In the U.S., scientific names are called United States Adopted Names (USAN), and are assigned by the USAN Council, which consists of the American Medical Association, the United States Pharmacopeial Convention, and the American Pharmacists Association. In the U.K., they are called British Approved Names (BAN names), and the World Health Organization assigns International Nonproprietary Names (INN names) in English, Latin, French, Russian, Spanish, Arabic, and Chinese. For example, Tylenol's USAN is "acetaminophen," but its INN and BAN is "paracetamol." (See Figure 3 on page 24).

Preserving Nuance

Typically, there are many synonyms for a chemical name in both the source and target languages. Choosing which synonym is most appropriate depends upon the context, which introduces a degree of subtlety, tone, and meaning to a translation. For some tasks, the goal may be to bring the document up to date by using the names currently recommended for a particular field. For other tasks, the goal may be to render a translation that is faithful to the original, thereby preserving the flavor of the text (akin to preserving "thou dost" in Shakespearean literature).

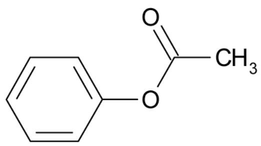
As an example, the names "acetic acid" and "ethanoic acid" both describe the same molecule, "CC (=O)O," the first being the commonly used traditional name, and the second being the systematic name. This distinction is preserved in many languages, including German (*Essigsäure* versus *Ethansäure*); Dutch (*azijnzuur* versus *ethaanzuur*); Polish (k*was octowy* versus *kwas etanowy*); Japanese (酢酸 versus エタン酸); and Chinese (醋酸 versus 乙酸).

The challenge is when one form is preferred over another nationally. For example, the traditional name is more common in English, whereas the systematic name is more prevalent in Chinese. Some languages fail to preserve these distinctions, making the translations from one to the other lose some of their meaning. For example, the element "mercury" in English has the archaic common name of "quicksilver," which can be preserved in Spanish as azogue rather than mercurio, but it is a distinction that is lost in many languages, such as German, where both become quecksilber. To return to our Tylenol example, the more up-to-date way to write this in Chinese would be "N-(4-羟基苯基)乙酰胺," but "N-(对羟基苯基) 乙酰胺" is the more traditional name.

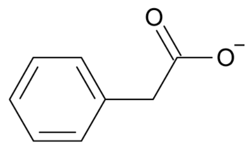
Subtle changes in hyphenation, spacing, and capitalization can denote completely different compounds.

One significant cause for the differences in chemical nomenclature forms among languages is the inevitable delay in translation and adoption of IUPAC standards. As the preferred chemical naming standards continue to evolve in English, there is an unfortunate lag with their implementation and adoption in other languages. This leads to some translations appearing dated and quaint. For example, the Chinese form of "methyl ben-

Figure 4

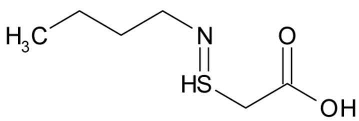


Phenyl acetate

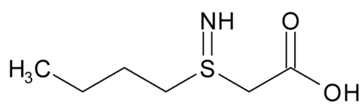


Phenylacetate

Figure 5

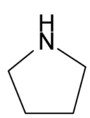


N-butylsulfinimidoylacetic acid

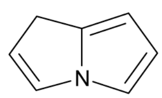


n-butylsulfinimidoylacetic acid

Figure 6



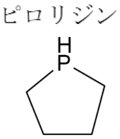
Pyrrolidine
ピロリジン



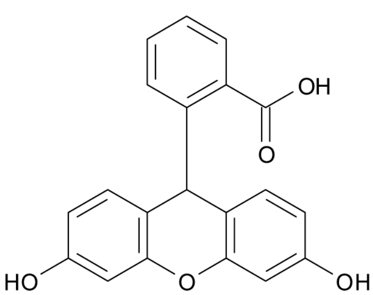
Pyrrolizine
ピロリジン

$PH_5^{(V)}$

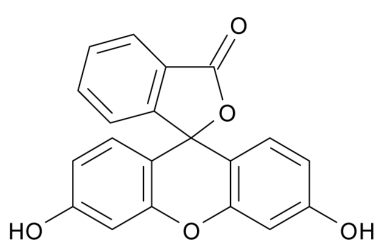
Phosphorane
ホスホラン



Phospholane
ホスホラン



Fluorescin
フルオレスセイン



Fluorescein
フルオレスセイン

zoate" is 苯甲酸甲酯, which is literally a translation of the more traditional "benzoic acid methyl ester."

A final subtle example of nuance in chemical nomenclature is the effect of dialect. Even in the English language, British, Canadian, and Australian chemical names differ from American chemical names. For example, the list below shows the names of elements in traditional British versus American spellings. The International entry describes the "International English" IUPAC-preferred spelling, which is required for world patent filings.

- **American:** Sulfur, Aluminum, Cesium
- **British:** Sulphur, Aluminium, Caesium
- **International:** Sulfur, Aluminium, Caesium

Examples of Complications

The near infinite number of possible molecules, each potentially with several names, limits the traditional use of dictionaries. Whilst technical lexicons can be helpful for commonly occurring substances, they are of limited use in "composition of matter" chemical patents, which by definition must describe a novel compound. Each such patent can frequently claim many millions of novel chemical structures, making it exceptionally unlikely that any names it contains would previously have been translated. This same issue also limits the utility of software based on translation memories, where chemical names are rarely reused in many documents.

A significant source of complications in translating chemical nomenclature is that the lexical structure and morphology of IUPAC compound names are often very different from those of the source language. For example, words are typically broken by spaces or punctuation in an English text. In IUPAC names, separate lexemes, such as "acet" and "amide" in "acetamide," are not separated by spaces, and full names often contain spaces, punctuation, numerals, brackets, superscripts, and even Greek characters. Subtle changes in hyphenation, spacing, and capitalization can denote completely different compounds. Typically, translation memory-based software will need special rules to correctly identify the start and end of each name.

An example of the significance of spacing is that the chemical name "phenyl acetate" denotes a different molecule from that specified by "phenylacetate." (See Figure 4 on page 26.)

An example of the significance of capitalization is that "N-butylsulfinimidoylacetic acid" represents a different chemical structure than "n-butylsulfinimidoylacetic acid." (See Figure 5 on page 26.)

An example of an interesting minor technical annoyance is the Chinese use of symbols used to denote elements. In Chinese, every element has its own character, requiring new symbols to be added to the character set as new heavy elements are discovered. This moving target means that elements over atomic number 104, such as seaborgium, fail to be displayed correctly in Web browsers or to be handled correctly in translation software. The reason: traditional Chinese characters were added only to version 3.1 of the Unicode standard (2001), and simplified Chinese characters have yet to be added (as of the most recent version 4.1 of the Unicode standard, 2005).

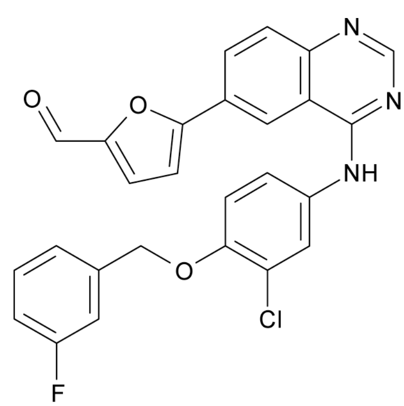
A potentially more serious cause for concern is that some languages are unable to express real distinctions found in chemical compounds in English, resulting in unavoidable ambiguities. Examples of this in Japanese are shown in Figure 6 on page 26.

Due to the lack of regularity, common names must be translated via a dictionary.

Obfuscation of Chemical Patents

Usually, the task of a translator is to produce a rendition that a native speaker can understand easily. Perhaps a unique exception to this rule is the task of translating patents. A patent is an exclusive monopoly granted by the state to an inventor (or his assignee) for a limited time in exchange for disclosure of an invention. Although patents require disclosure in sufficient detail for a person skilled in the art, there is no obligation to make the task easy or to limit the amount of "skill" required. In chemistry, this practice dates back to medieval alchemists, who first introduced chemical nomenclature to hide their work and its results from lay folk and competing alchemists. It is ➡

Figure 7



Original Japanese:
5-(4-[3-クロロ -4-(3-フルオロベンジルオキシ)-アニリノ]-6-キナゾリニル)
-フラン -2-カルバルデヒド

Official English Translation:
5. –(4-[aniline [the 3-chloro- 4 –(3-fluoro benzyloxy)-]]-6- chinae-cortex ZORINIRU)- franc 2-carbaldehyde

Correct English Translation:
5-(4-[3-chloro-4-(3-fluorobenzyloxy)-anilino]-6-quinazolinyl)-furan-2-carbaldehyde

not unusual for pharmaceutical patents to use extremely broad and generic titles, such as "therapeutic compounds," making the titles alone almost useless for searching and indexing. Pharmaceutical patents also take advantage of the ability to assign several different names to a single compound or molecule. There are examples of a single patent document using four or five different names for the same compound, none of which are the official IUPAC-preferred name.

The use of obfuscation in patent filings leads to two goals of translation: 1) the translation of competitor patents so that they are understandable to the client; and 2) the translation of client patents so that they are difficult for competitors to understand. For example, under the European Patent Convention, European patents may be filed in one of several languages, and the filing language may be chosen based on the ease with which it can be translated and understood by a competitor.

Resources for Translators

So what are some resources for terminology for translators and interpreters working in this field? Space restrictions do not permit me to list them all, but here are a few to get you started. (Make sure to check out additional publications and websites listed at the end of this article.)

IUPAC: An extremely useful resource for translators is the official national translation of IUPAC's naming recommendations. Once an IUPAC standard is ratified internationally, it may officially be translated into other languages. A list of current translations may be found at IUPAC's Nomenclature Books website, www.chem.qmul.ac.uk/iupac/bibliog/books.html.

Another great online resource is the current draft of IUPAC's provisional recommendations for organic nomenclature, which describe the latest standards for naming molecules in English, available at http://old.iupac.org/reports/provisional/abstract04/BB-prs310305/CompleteDraft.pdf. One convenient feature of these translations is that once a chemistry translator is familiar with one version, finding the equivalent section in another language is much easier.

Dieter Hellwinkel: Hellwinkel's excellent textbook on chemical naming, *Systematic Nomenclature of Organic Chemistry: A Directory to Comprehension and Application of its Basic Principles,* is available in both English and German versions.

Related Links

American Chemical Society
www.acs.org

American Medical Association
www.ama-assn.org

American Pharmacists Association
www.aphafoundation.org/home/index.cfm

Chemical Abstracts Service
www.cas.org

European Patent Convention
www.epo.org/patents/law/legal-texts/epc.html

International Union of Pure and Applied Chemistry
www.iupac.org

International Union of Pure and Applied Chemistry Nomenclature Books
www.chem.qmul.ac.uk/iupac/bibliog/books.html.

United States Pharmacopeial Convention
www.usp.org

World Health Organization
www.who.int/en

Wikipedia: Another excellent resource for information is Wikipedia and its various translated forms. Many frequently occurring molecules of industrial or scientific interest now have their own Wikipedia pages, and the translations of these pages provide an unparalleled resource.

Translators need to be cautious when using online resources. Unfortunately, the current state-of the-art in machine translation is frequently tripped up by the unique properties of chemical nomenclature. The SYSTRAN system, used by Google and Yahoo! translation services, currently translates the Japanese systematic name for caffeine, from "1,3,7- トリメチルプリン -2,6- ジオン" as "1,3,7- 2,6- trimethyl pudding – dione" instead of "1,37-trimethylpurine-2,6-dione." Even the English translations provided by the Japanese patent office are nonsensical, as demonstrated by the example in Figure 7 on page 27 for Japanese Patent Number 2008-50363.

Whilst the situation is beginning to improve with the first generation of software tools specializing in the translation of chemical nomenclature, the need to preserve nuance as described in this article, will restrict the utility of these software packages to the area of machine-assisted translation.

References

Bünzli-Trepp, Ursula. *Systematic Nomenclature of Organic, Organometallic, and Coordination Chemistry: Chemical-Abstracts Guidelines with IUPAC Recommendations and Many Trivial Names* (EPFL Press, 2007).

Fox, Robert B., and Warren H. Powell. *Nomenclature of Organic Compounds: Principles and Practice, 2nd Edition* (Oxford University Press, 2001).

Hellwinkel, Dieter. *Die Systematische Nomenklatur der Organischen Chemie: Eine Gebrauchsanweisung* (Springer-Verlag, 2006).

Hellwinkel, Dieter. *Systematic Nomenclature of Organic Chemistry: A Directory to Comprehension and Application of its Basic Principles* (Springer-Verlag, 2001).

International Union of Pure and Applied Chemistry. *Nomenclature of Organic Chemistry: Sections A, B, C, D, E, F, and H.* Edited by Rigaudy, J., and S. P. Klesney (Pergamon Press, 1979).

International Union of Pure and Applied Chemistry. *A Guide to IUPAC Nomenclature of Organic Compounds, Recommendations 1993.* (Blackwell Scientific, 1993).

Mitchell, A. D *British Chemical Nomenclature* (Edward Arnold & Co. publishers, 1948).

Naming and Indexing of Chemical Substances for Chemical Abstracts. Appendix IV of CA Index Guide (Chemical Abstracts Service, 2007).

Nyitrai, J., and J. Nagy. *Útmutató a szerves vegyületek IUPAC-nevezéktanához* (Magyar Kémikusok Egyesülete: Budapest, Hungary, 1998).

Peterson, W. R. *Formulacion Y Nomenclatura Quimica Organica* (Edunsa: Barcelona, Spain, 1993).

Polskie Towarzystwo Chemiczne. *Nomenklatura Związków Organicznych* (Państwowe Wydawnictwo Naukowe: Warsaw, Poland, 1992).

Polskie Towarzystwo Chemiczne. *Przewodnik Do Nomenklatury Związków Organicznych* (Narodowy Komitet Międzynarodowej Unii Chemii Czystej I Stosowanej: Warsaw, Poland, 1994).

Rechtsvorshriften für Gefährliche Stoffe: Einstufung und Kennzeichnung in der Europäischen Gemeinschaft. 67/548/EWG (Commission of the European Communities, 1987).

Sayle, Roger. "Foreign Language Translation of Chemical Nomenclature by Computer," *Journal of Chemical Information and Modeling* (February 2009, Vol. 49, No. 4), 519-530, February 2009. http://pubs.acs.org/doi/abs/10.1021/ci800243w

Wikman, S. *Organisk-kemisk Nomenklatur* (Studentlitteratur, 2004).

ata

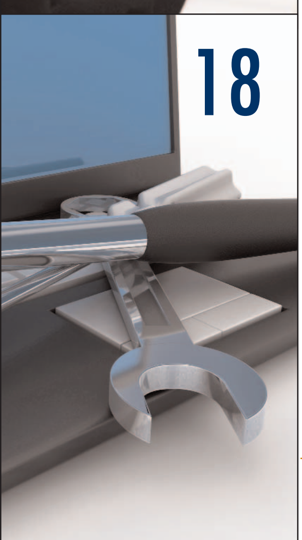
October 2009
Volume XXXVIII
Number 10

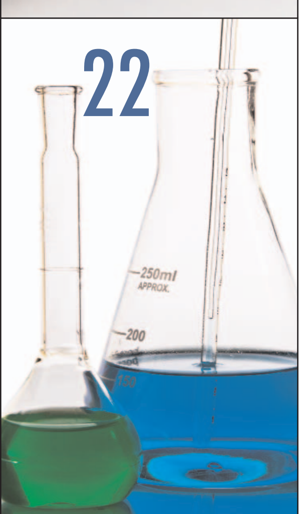
A Publication
of the
American
Translators
Association

Contents | October 2009

American Translators Association
225 Reinekers Lane, Suite 590 • Alexandria VA 22314 USA
Tel: +1-703-683-6100 • Fax: +1-703-683-6122
E-mail: Chronicle@atanet.org • Website: www.atanet.org

50 Years
1959 – 2009





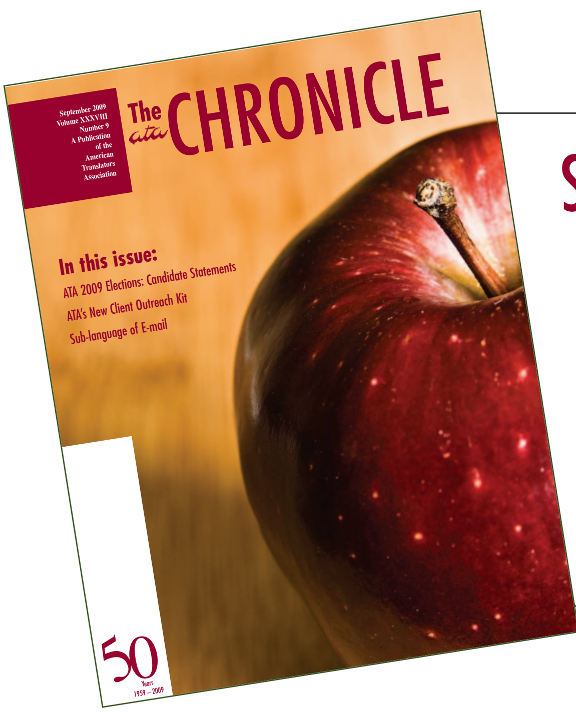
Columns and Departments

Lillian Clementi is ATA's School Outreach Coordinator and a member of ATA's Public Relations Committee. As managing principal of Lingua Legal, a translation practice based in Arlington, Virginia, she translates from French and German into English, specializing in law and business. Contact: Lillian@LinguaLegal.com.

Diane Howard is a freelance translator specializing in medical and pharmaceutical translation from Chinese and Japanese into English. She is an ATA-certified Japanese→English translator, and has taught translation for the University of Chicago Graham School and the University of Denver. She holds certificates in medical writing and editing and in clinical trials management from the University of Chicago and in technical Japanese studies from the University of Wisconsin-Madison. She is currently a PhD student in translation and intercultural studies at the Universitat Rovira i Virgili, Tarragona, Spain, and is writing a dissertation on translation testing. Contact: diane.howard@worldnet.att.net.

Roger Sayle is currently the vice-president at OpenEye scientific software, which develops software to assist drug discovery in the pharmaceutical industry. He is an expert in chemical nomenclature. In addition to ATA, he is a member of the American Chemical Society and the International Union of Pure and Applied Chemists. He is author of OpenEye's Lexichem software for automatically generating names for chemical structures and for extracting structures from names. Recently, he has been working on machine-assisted translation of chemical nomenclature among 18 languages. He received his PhD from the University of Edinburgh, U.K., after which he spent several years conducting drug discovery research at GlaxoWellcome before moving to the U.S. Contact: roger@eyesopen.com.

Jost O. Zetzsche is an ATA-certified English→German translator and a localization and translation consultant. A native of Hamburg, Germany, he earned a PhD in the field of Chinese translation history and linguistics, and began working in localization and technical translation in 1997. In 1999, he co-founded International Writers' Group and, in 2008, TranslatorsTraining.com. He currently sends out a free biweekly technical newsletter for translators (www.internationalwriters. com/toolkit). Contact: jzetzsche@internationalwriters.com.



Send a Complimentary Copy

If you enjoyed reading this issue of *The ATA Chronicle* and think a colleague or organization would enjoy it too, we'll send a free copy.

Simply e-mail the recipient's name and address to Maggie Rowe at ATA Headquarters—maggie@atanet.org—and she will send the magazine with a note indicating that the copy is being sent with your compliments.

Help spread the word about ATA!