Word Formation and Transparency in Medical English

Word Formation and Transparency in Medical English

Edited by

Pius ten Hacken and Renáta Panocová

Cambridge Scholars Publishing

Word Formation and Transparency in Medical English

Edited by Pius ten Hacken and Renáta Panocová

This book first published 2015

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-4438-8002-7
ISBN (13): 978-1-4438-8002-2

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

MEDICAL LANGUAGE, WORD FORMATION AND TRANSPARENCY

PIUS TEN HACKEN AND RENÁTA PANOCOVÁ

As a side effect of the rapid progress in medical research and of the emergence of new medical conditions, medicine is a domain where new concepts have to be named more frequently than in many other domains. Because of the prominent position of English in medical research, most of these new concepts are first named in English. This volume takes this situation as a background for the study of naming strategies used for such new concepts. Before introducing the individual chapters, in this introduction we will present some general thoughts about the nature of medical language, the role of word formation in naming, and the question of what constitutes transparency.

1 Medical language

A first question to be asked in a volume devoted to medical language is what kind of language is designated by the expression *medical language*. In English, we use the adjective *English* as a noun when it refers to the language. In Slavic languages, this is not common. Thus, in Polish one rather finds *język angielski* ('language English', i.e. the English language) than just *angielski* ('English') to refer to the language. However, for specialized languages, we cannot leave out the noun in English. We cannot use *medical* to mean *medical language*. This provides a first indication that *medical language* is not perceived as a language in the same way as *English*.

For a long time, linguists assumed that English, Dutch, Slovak, and other languages should be taken as the basic objects of study in linguistics. In 19th century historical-comparative linguistics, the historical development of such objects and their relationships were taken as central.

Thus, August Schleicher (1821-1868) proposed a genealogical tree of Indo-European languages (cf. Collinge 1995: 198). With his distinction between synchronic and diachronic linguistics, Saussure (1916) changes the emphasis, focusing more on the state of such languages at a particular point in time than on their development and relations, but he maintains the idea of a language as an object of study. A question that was not systematically pursued was how we can establish whether a language in fact has a particular property. What is the empirical basis for the claims that English has no grammatical gender, Dutch has a phonemic contrast between /f/ and /v/, and Slovak has seven nominal cases?

A central insight into the nature of language is Chomsky's (1965: 4) distinction between *competence* and *performance*. As explained in more detail in ten Hacken (2007: 42-46), both competence and performance are empirical objects, the former the knowledge of language of an individual speaker as realized in the speaker's brain, the latter the utterances and texts produced applying this knowledge. As argued by ten Hacken (2007: 274-281), it was only in the course of the 1970s that Chomsky realized that in this model there was no place for languages such as English. Some of the properties that we ascribe to English are clearly not compatible with competence and performance. Thus, English is the language of the UK, the USA, and a number of other countries. It is one of the drafting languages of the EU. It is the language of Chaucer, Shakespeare, Jane Austen and J.K. Rowling. It has been spoken since the 6th century CE. In the sense in which *English* has such properties, it is not an empirical object. We can only arrive at a notion of *English* by classifying linguistic knowledge (i.e. speakers) or linguistic output (i.e. texts and utterances).

This insight changes the way in which a claim such as the one that English has no grammatical gender can be tested. We can determine whether this claim holds for English as competence only to the extent that we narrow down the scope of the claim to an individual speaker. For English as performance, we can test the claim for a particular corpus of texts and utterances. For English as a language, it is not possible even in principle to collect direct empirical evidence. We can only approach the question either by first determining that a speaker or a set of speakers is characteristic of English and then study their competence or by first collecting a corpus to which we assign the label that it is representative of English and study its properties. In both cases, we take decisions that are not determined by the data before we can do any empirical investigation. English is not an empirical object.

Medical language differs from English in several ways. It is not a natural question, for instance, to ask whether someone speaks medical

language, as opposed to the question whether someone speaks English. Medical language is also language-specific. English medical language differs from Dutch or Slovak medical language. This raises the question of the relationship between English medical language and English, but also of the relationship with Dutch and Slovak medical language.

A framework that has often been used as a basis for approaching such questions is that of sublanguages. Here, the idea is that English medical language is a sublanguage of English. The original idea of sublanguages stems from Harris (1968). Kittredge (1987: 59-60) defines a sublanguage L_S as a subsystem of a language L, such that L_S is part of L, but has a more restricted domain and community of speakers. L_S is a consistent and complete linguistic system, so that it has its own sets of vocabulary items and syntax rules. Kittredge proposes to derive them from the analysis of a corpus. It is fairly straightforward to do this for the vocabulary. For the grammar rules, Kittredge (1987: 62-63) makes the following observations:

> First, in a sublanguage, the rules for constructing meaningful sentences can be made much more precise than in the language as a whole. These rules can be related in terms of word classes which are discovered by studying the distributional properties of words in texts. Second, in a sublanguage the rules for constructing sentences may be quite different from (and even contrary to) the rules for sentences in the 'standard' language.

For medical English, this means that we have to collect a corpus of texts as a preliminary. In compiling such a corpus, we have to take decisions. The corpus is of course an empirical object, but the decision whether a text belongs to medical English is based on a judgement, not on any empirical fact. A crucial question seems to be whether the corpus as a whole is representative of English medical language. However, there is no way to go beyond intuitive judgements for answering this question.

In our view, it would be misguided to conclude from this argument that the concept of *medical English* should not be used. Otherwise we would not edit a volume that has *medical English* in its title. An important observation in this respect is that not all concepts we evoke in an academic text have to be precisely delimited terms. Although Chomsky often made the point that there is no object called *English*, Chomsky & Lasnik (1995: 33) state that "[i]n English, generally only objective Case-assigning verbs can occur in the passive". The use of *English* in this quotation is not incoherent with the observation that there is no such object. In this quotation, English is used pretheoretically.

In the same way we can make statements about medical English, without implying that there is an entity called *medical English* for which

the statement is correct. However, we cannot determine, for example, the exact number of words in medical English. In this volume, *medical English* will be used as a pretheoretical notion. We do not make any claims that depend on the precise boundaries of the concept.

2 Word formation

Word formation is a system of rules that can produce new words on the basis of existing lexical items. Word formation can be distinguished from syntax. Both take lexical items as their input, but whereas syntax produces sentences to express thoughts, word formation produces words to name concepts. Word formation has an onomasiological function and changes the lexicon.

In medical language, word formation is particularly prominent because there is a steady growth in the number of concepts that need to be named. For the study of word formation this naming need is important, because it is a decisive factor in activating word formation. Only when there is a new concept that needs a name will word formation rules be activated. However, naming needs can also be fulfilled in other ways. The most prominent alternative naming procedures are sense extension and borrowing.

An example of sense extension in medical English is the use of *cell* for a small unit of the body. The original meaning was a small room in a monastery or prison. This example shows how sense extension is based on metaphoric or metonymic sense relationships. It also illustrates the notion of *onomasiological coercion*. Much of the meaning of the resulting term is determined by the concept we started with, independently of the naming mechanism and the input. That a cell has a nucleus and multiplies by fission cannot be derived from *cell* in the sense of a small room.

Given the large degree of international exchange in the field of medicine, it is not surprising that borrowing plays an important role as a naming mechanism. At the level of research, where new concepts are discovered and named, Latin used to be the language of international communication. Often, also Ancient Greek words were used in their Latinized form. By the time medicine went through the transition from a craft-like practice to an applied science, which as Bynum (1994) argues occurred in the 19th century, most medical research was no longer published in Latin. However, for the naming of new concepts Latin and Greek continued to be used in the form of neoclassical word formation. Especially at the time when several major European languages were used for international scientific communication, the use of neoclassical

formations that could be recognized equally in each of these languages was a useful aid to understanding.

In the current situation, the overwhelming majority of medical research is published in English. As opposed to fields such as astronomy or mathematics, however, for communication in the field of medicine there is much more pressure to accommodate a wide range of other languages. This is because in medicine there is always a need to communicate in vernacular languages of the communities. Medical communication is not limited to researchers, but also includes patients and their relatives. At some point in the chain between researcher and patient, the terminology has to be translated. Here the question arises whether an English name should be borrowed or a language-specific alternative naming device used instead.

It is in this configuration that word formation operates. As opposed to alternative naming mechanisms, word formation is based on the application of rules. At the same time, in common with other naming mechanisms and in contrast to syntax, word formation is not rule-driven. The starting point is not a form and meaning determined by a word formation rule, but the need to name a specific concept. Word formation rules constrain the meaning of their output, but it is in general not fruitful to start from the word formation rule in trying to explain the full meaning of the resulting word. As argued by ten Hacken (2013), a much more promising approach is to start from the concept to be named. On the basis of this concept, a word formation rule and an input to this rule are chosen, but the full meaning only arises through onomasiological coercion. This process can be illustrated on the basis of (1).

(1) cuvette oximeter

In (1), we have a compound consisting of the head *oximeter* and the non-head *cuvette*. The head looks like a neoclassical word formation, but in fact, *oxi* stands for *oxygen*, not for the Ancient Greek ὀξύς [oxys] (‘sharp’). An oximeter is an instrument for measuring the quantity of oxygen in blood. The non-head is a loanword from French. At the point where (1) is formed, *oximeter* and *cuvette* are words in the lexicon. The meaning of a compound is underspecified. We can deduce from the form that (1) designates a kind of oximeter that is related to (a) cuvette, but in order to understand the meaning in more detail we have to know the concept for which it was coined as a name. Stedman (1990) gives the meaning as in (2).

(2) an o[ximeter] that reads the percentage of oxygen saturation of the blood as it passes through a cuvette outside the body

The definition in (2) is given in a run-on entry to *oximeter*, which explains the abbreviation of this word. Significantly, the definition is almost entirely a description of the relation between the head and the non-head in (1).

Word formation rules are used to produce new words. This is not restricted to the speaker who uses this word for the first time in the language. First of all, it is hardly possible to determine which speaker used (1) for the first time. Secondly, the notion of *language* in "for the first time in the language" is the non-empirical, pretheoretical sense of language, of which we cannot determine the precise boundaries. However, every speaker coming across (1) and not having this word in their mental lexicon will use a word formation rule to interpret it. Depending on the speaker's needs, the word can then be stored in their lexicon or not. The meaning associated with (1) depends on the speaker's knowledge of and experience with the concept.

3 Transparency

There are different concepts that can be used to describe the relationship between the form and meaning of words. *Transparency* should be distinguished from *motivation* and *iconicity*. All of them contrast with Saussure's (1916: 100) statement in (3).

(3) Le lien unissant le signifiant au signifié est arbitraire[1]

Saussure gives the example of French *bœuf* ('ox') and its German translation *Ochs*. If (3) did not hold, we would have to find an explanation why not all languages have the same word for the same concept. Saussure (1916: 101) clarifies that *arbitraire* in (3) does not mean that any speaker can choose a *signifiant* at will, but that the form is "*immotivé*" ('unmotivated'), i.e. there is no natural link between the form and the meaning.

Whereas the existence of onomatopoeia is at best a marginal counterexample to (3), Saussure (1916: 180-184) modifies the scope of (3) somewhat in view of morphological relationships. He distinguishes

[1] 'The link uniting the signifier (i.e. the form) to the signified (i.e. the meaning) is arbitrary' [our translation, PtH & RP]

absolute and relative arbitrariness, so that, for instance, *happy* and *sad* are fully arbitrary, but *unhappy* is only partially arbitrary.

As indicated by the use of *immotivé*, the opposite of *arbitrary* for Saussure is rather *motivated* than *transparent*. In addition, we have the term *iconicity*, used in a similar way. However, there are different shades of meaning involved and it is worth distinguishing them. In explaining the contrast, we use the compound (1) as an example. The degree of motivation concerns the extent to which the speaker is guided to use this expression for the instrument it refers to. One of the decisions involved is the one to use a compound.

The degree of transparency concerns the extent to which the reader or hearer is helped by the form in the task of determining the meaning. Given that compounds in English are regularly right-headed, a reader will understand (1) as the name of a kind of oximeter. In this sense, compounding makes (1) more transparent than a non-compound might be, in particular a simple expression that is not the result of word formation. The underspecification of the relation between the head and non-head, however, reduces the transparency.

In the discussion of iconicity, neither the role of the speaker nor of the hearer is taken into account. As Dressler (2005: 268) states, the concept of *icon* is based on work by Charles S. Peirce (1839-1914). In the context of natural morphology, Mayerthaler (1981: 23) introduces *konstruktioneller Ikonismus* ('constructional iconicity'). The general idea is that more complex concepts have longer names. This applies to (1) in the sense that *cuvette oximeter* refers to a more specific type of instrument than *oximeter*.

In natural morphology, iconicity is connected to two related concepts, diagrammaticity and biuniqueness. Dressler (2005: 269) gives compounds as a typical example of diagrammaticity, because the semantic head is also the morphosyntactic head. A compound such as (1) is diagrammatic to the extent that its head *oximeter* determines at the same time the meaning, in the sense that (1) designates a type of oximeter, and the syntactic properties, to the extent that (1) is a countable noun like *oximeter*. Whereas iconicity is more of a quantitative notion, based on the amount of information, diagrammaticity concerns the relative contribution of morphological elements.

By biuniqueness, the one-to-one correspondence between form and meaning is meant. As noted by Dressler (2005: 274), this is particularly important in terminology. In classical approaches such as Wüster (1991), biuniqueness is almost axiomatic. However, as Dressler notes, it is only aimed for on a domain-internal basis. The fact that *induction* is used in medicine for the artificial stimulation of child birth but in mathematics for

a particular type of proof is not problematic. Together, iconicity, diagrammaticity and biuniqueness can be used to describe the way in which meaning and form relate to each other without referring to a speaker or a hearer.

The three terms of *motivation*, *transparency*, and *iconicity* have not always been distinguished consistently in the literature. However, as far as we can judge, where two or three of them have been used in a contrastive sense, *motivation* is usually connected to the speaker perspective, *transparency* to the hearer perspective, and *iconicity* to a perspective that focuses on words as abstract objects. For *iconicity*, this means that it is independent of competence and performance, which in our view makes it a less interesting property for the study of the role and effects of word formation in medical English. The question of motivation is especially relevant in studies of productivity, an issue that is not central in this volume. It is therefore natural here, in our opinion, to make transparency the focus of the study of how word formation interacts with the relationship between form and meaning.

4 Overview of chapters

Our volume consists of eight chapters that can be divided into two parts of four chapters each. The first part concentrates on the study of transparency from a monolingual perspective; the second part contains studies of translation.

In chapter 1, Rachel Bryan focuses on International Nonproprietary Names (INNs) for pharmaceutical substances. In this context, transparency has a particular significance, because the names of the substances are used to identify the correct medical treatment for particular patients. Therefore, the World Health Organization (WHO) has elaborated a set of guidelines for arriving at an INN that is at the same time maximally transparent and sufficiently distinct. Bryan investigates how these guidelines are used in practice and to what extent they achieve their goals. One of the special features of INNs is that they are composed of formatives that have been assigned a meaning rather than on a system of word formation that has emerged naturally in a language.

In chapter 2, by Pilar León-Araúz, we turn to the psychiatric domain. She investigates the correlation between terminological variation and transparency. In the psychiatric domain, there are many synonyms that are distinguished in various ways. On one hand, there is a variety of registers ranging from formal to colloquial. These registers often correlate with the type of participants in the oral or written communication. On the other

hand there are connotative factors that may make certain expressions more accessible or less acceptable. León-Araúz describes how a database to account for this variation is structured and gives some results that have been found by analysing the occurrence of variants in a specially constructed corpus.

In chapter 3, Pius ten Hacken takes an approach that can be characterized as corpus-based and semasiological in the sense that he studies the medical terminology used in a single text. This text is Gersdorff & Gérard's (2011) *Atlas of Middle Ear Surgery*, an introduction to a specialized area of medicine, written for otologists. Ten Hacken classifies the terms extracted from this text along morphological criteria and analyses how properties of the word formation rules involved affect the transparency of the terms. In his collection of terms, there is a large preponderance of compounds.

Chapter 4, by Renáta Panocová, delimits the domain of study more in a morphological way than on the basis of the medical specialization. As observed above, neoclassical word formation is an important source of terms in the medical domain. Even though nowadays few medical researchers have a sufficient command of Ancient Greek and Latin to write texts in these languages, neoclassical elements are still commonly used to name new medical concepts. Panocová compares the transparency of such formations with the transparency of some commonly used alternatives, in particular eponyms and abbreviations.

With chapter 5, we enter the second part of the volume, with studies involving translation. This chapter, written by Nina Patton, María Fernández Parra and Rocío Pérez Tattam, delimits the domain in morphological terms, looking at nominal compounds. There are two language pairs involved, English-Spanish and English-Slovak. A major challenge of these language pairs is that Spanish and Slovak do not have an immediately corresponding construction for English compounds. The authors investigate which constructions are used in the translation of English compounds in each language and to what extent the choice of construction is influenced by the relation between the head and the non-head of the English compound.

In chapter 6, Sevda Pekcoşkun takes a different perspective of translation. Her language pair is English-Turkish and her first question is which translation strategies are used in the translation of popular medical texts. Whereas for research articles in medicine, we can expect that many readers will put up with an English version in which they recognize many of the terms immediately as borrowings or internationalisms, for popular medical texts, it is important that Turkish speakers can read and

understand them. This leads to Pekcoşkun's second question, where the influence of the choice of translation strategy on the intelligibility for the target readership is investigated. She uses questionnaires, asking readers to compare how easy different translations are for understanding.

The last two chapters study the language pair English-Polish. In chapter 7, by Mariusz Górnicz, the perspective is that of medical specialists and how they render English terminology in Polish. Górnicz argues that there is a strong resistance not only to borrowing English terms in Polish, but also to adopting the same structures in Polish as in English. A central concept in this context is what he calls *compression*, i.e. the techniques that lead to a more concise expression, so that terms are more efficient in communication than full descriptive phrases.

Finally, in chapter 8, Szymon Machowski returns once more to compounds. His research focuses on the domain of infectious diseases and he applies a classification based on two independent features. On one hand, he introduces four semantic categories and on the other, four formal properties of the expression. He then considers how English terms in these different categories are translated into Polish.

Most contributions to this volume are based on presentations at the Seminar 'Word formation and transparency in Medical English', organized by the editors at the 12th Conference of the European Society for the Study of English (ESSE) in Košice. We would like to thank all participants to this seminar who gave comments and asked questions, triggering a lively discussion that led to a better formulation of the ideas presented. We would also like to thank Sam Baker of Cambridge Scholars Publishing for suggesting the idea to edit a volume based on this workshop. In the production of this volume, we were helped by the very responsive authors who helped us to realize this idea. In producing the manuscript, we benefited from editorial support by Christina Muigg and proofreading by David Galvin.

We hope that the volume will be useful for medical text writers and translators in that it offers a range of perspectives on problems that they have to solve every day. At the same time, terminologists will find here a number of case studies. Morphologists, especially those working on word formation, may benefit from the study of naming practices in a number of areas that are related in the sense that they are all in the field of medicine, yet quite different in their actual approach to naming.

References

Bynum, W.F. (1994), *Science and the Practice of Medicine in the Nineteenth Century*, Cambridge: Cambridge University Press.

Chomsky, Noam (1965), *Aspects of the Theory of Syntax*, Cambridge (Mass.): MIT Press.

Chomsky, Noam & Lasnik, Howard (1995), 'The Theory of Principles and Parameters', in Chomsky, Noam, (ed.), *The Minimalist Program*, Cambridge, MA: MIT Press, pp. 13-127.

Collinge, N.E. (1995), 'History of Comparative Linguistics', in Koerner & Asher (eds.), *Concise history of the language sciences from the Sumerians to the cognitivists*, Oxford, pp. 195-202.

Dressler, Wolfgang U. (2005), 'Word-Formation in Natural Morphology', in Štekauer, Pavol & Lieber, Rochelle (eds.), *Handbook of Word-Formation*, Dordrecht: Springer, pp. 267-284.

Gersdorff, Michel & Gérard, Jean-Marc (2011), *Atlas of Middle Ear Surgery*, Stuttgart: Thieme.

ten Hacken, Pius (2007), *Chomskyan Linguistics and its Competitors*, London: Equinox.

—. (2013), 'Semiproductivity and the place of word formation in grammar', in ten Hacken, Pius & Thomas, Claire (eds.), *The Semantics of Word Formation and Lexicalization*, Edinburgh: Edinburgh University Press, pp. 28-44.

Harris, Zellig (1968), *Mathematical Structures of Language*, New York: Interscience.

Kittredge, Richard I. (1987), 'The significance of sublanguage for automatic translation', in Nirenburg, Sergei (ed.), *Machine Translation: Theoretical and Methodological Issues*, Cambridge: Cambridge University Press, pp. 59-67.

Mayerthaler, Willi (1981), *Morphologische Natürlichkeit*, Wiesbaden: Athenaion.

Saussure, Ferdinand de (1916), *Cours de linguistique générale*, Charles Bally & Albert Sechehaye (eds.), Édition critique préparée par Tullio de Mauro, Paris: Payot, 1981.

Stedman (1990), *Stedman's Medical Dictionary*, 25th edition, Hensyl, William R. (ed.), Baltimore: Williams & Wilkins.

Wüster, Eugen (1991), *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*, Bonn: Romanistischer Verlag.

CHAPTER ONE

TAXONOMY AND TRANSPARENCY IN INTERNATIONAL PHARMACEUTICAL NOMENCLATURE

RACHEL BRYAN

The language of medicine, although highly specialised, has a broad usership comprising multiple strata of the population with varying levels of knowledge for multiple purposes. This usership includes general practitioners, consultants, nurses, pharmacists, patients, parents and caregivers. No single person holds a comprehensive knowledge of every area and so there is great variation in understanding of the terminology and the degree to which its use is specialised. Medication names such as *morphine*, *Benadryl*, *paracetamol* and *adrenaline* surround us in our daily lives and are an important and under-researched area of terminology.

In antiquity, medications were named after the gods, e.g. morphine after Morpheus, the god of dreams and anandamide after Sanskrit *ananda*, 'bliss, delight' (OED). In the present day, pharmaceutical substances are named within a complex system of nomenclature which is managed by multiple government bodies. As illustrated in Figure 1-1, a pharmaceutical substance such as salbutamol (an asthma medication) will have three types of name.

- One chemical name, based upon the chemical formula of the substances, indicating the position of hydroxy groups, the length of the carbon chain and so on. This name is designated by the International Union of Pure and Applied Chemistry (IUPAC) and is published multilingually. There are some interesting translation problems in this area, but they are beyond the scope of this chapter.
- At least one brand, or proprietary name, chosen by the manufacturer that originally created the substance. This name is commercially driven, initially capitalized and legally bound to not imply any

therapeutic benefit. It is typically laconic and euphonious. Once out of patent (up to 20 years in the EU), the substance can be marketed by other companies and so will be assigned more brand names.

- At least one generic or non-proprietary name. On a global level, it will be assigned an International Nonproprietary Name (INN) by the World Health Organization (WHO), and in each country in which it is approved for use, it will be assigned a national generic name, such as a British Approved Name (BAN) in the UK, or a *Denominazione Comune Italiana* (DCIT) in Italy.

Chemical name	(*RS*)-4-[2-(*tert*-butylamino)-1-hydroxyethyl]-2-(hydroxymethyl)phenol formula: $C_{13}H_{21}NO_3$
Generic names	International Nonproprietary Name: salbutamol British Approved Name: salbutamol United States Adopted Name: albutamol
Proprietary names	Ventolin, Aerolin, Ventorlin, Asthalin, Proventil, ProAir

Fig. 1-1. The pharmacopoeial monograph for salbutamol[1]

1 International Nonproprietary Names

This chapter presents a qualitative analysis of the International Nonproprietary Name (INN) nomenclature, focusing in particular on the underlying conceptual taxonomy and semantic transparency. INNs will be the focus of this study as they are the most commonly used system of generic names, and their form is used by default in both the UK and the EU with only a few notable exceptions (Aronson 2000). There are over 8,000 INNs currently in use. INNs are designated by the WHO and are formally placed in the public domain to promote consistency of global communications between manufacturers, clinicians, prescribers and patients. The nomenclature is published in seven languages (WHO 1997). Given their international status, the name designation process in place must encompass a broad conceptual system and naming guidelines must be robust and stringently applied.

INNs are designated according to a set of guidelines (WHO 1997), which aim to achieve usability (pronounceable, legible, audibly perceptible, comprehensible and memorable), clarity (free from confusion)

[1] A pharmacopoeial monograph is a single document describing the name(s) and chemical formula of a pharmaceutical substance.

and taxonomy (showing relationship within the conceptual system). The WHO dictates that pharmacological relationship be shown by using a common 'stem', which may be a prefix, infix, suffix, or a 'freefix', and which can appear anywhere in the name. A 'stem' in this context is a word part to which a particular pharmacological meaning has been assigned and which is used to signify the relationship between substances. By using a common stem, substances are placed into pharmacological groups, related by anatomical target, therapeutic action, or chemical composition. The use of stems creates a taxonomic conceptual system for INNs and allows users to exploit this systematicity to increase retention, pronunciation and recognition of the names.

The INN programme began in 1952 and between 120 and 150 new names are designated each year. They are first created in Latin and this form is translated into the six official languages of the United Nations: English, French, Spanish, Russian, Chinese and Arabic. The Latin form of the name is used as the basis for translation into other European languages, such as Italian and Portuguese (Marecková *et al*. 2002).

Morphosemantic analysis of INNs is possible because their meaning is highly compositional, i.e. meaning is derived from the meanings of constituent parts (Deléger *et al*. 2009). In contrast to medical terminology in anatomy and general medicine, INNs are not full neoclassical compounds in that they cannot be parsed into elements directly derived from classical languages. INNs are composed of a random element, normally a prefix, and at least one stem. Stems are formed from three types of component. These types are listed in (1).

(1) a. abbreviations, such as the sub-stem *-tu-* in *situximab* denoting targeting tumorous tissue, or the stem *-kin* in *ilodecakin* denoting interleukin-type substances;
b. acronyms, such as the stem *-mab* in *urtoxazumab* denoting monoclonal antibodies; and
c. elements of chemical nomenclature. These can be seen as adapted neoclassical forms, such as the stem *-fos* (from Latin *phosphorous*) in *clofenvinfos*, denoting phosphorous derivatives.

2 Why is this important?

The World Health Organization (WHO) cites globalization, consumerism, growth in free markets, increased cross-border communication and the ubiquity of the Internet as agents of change in medicine and

pharmaceuticals, giving rise to new safety concerns. Furthermore, the increasingly global trade in pharmaceuticals and higher levels of regulatory complexity have impelled many intergovernmental organisations to make efforts towards harmonisation of regulatory activities to ensure consistent efficacy of pharmacovigilance efforts (WHO 2002).

Medication errors make up a high proportion of all patient safety events (Jordan & Kyriacos 2014; Ostini *et al*. 2012) and some result in overdose or adverse drug reactions, and can cause serious harm to patients (Aronson 2009); Runciman *et al*. 2003). Medication incidents in the UK resulted in 50 deaths between October 2011 and September 2012 (Jordan & Kyriacos 2014). It is estimated that medication errors cost the USA between $15bn and $28bn each year and that the USA spent an additional $213bn (8% of total healthcare spend) in 2012 on costs arising from medicines mismanagement, including medication errors (Aitken & Valkova 2013).

Medication errors may be a result of medicines having names that look alike or sound alike and are referred to as *LASA errors*. Examples of confused LASA pairs are given in (2).

(2) a. *mercaptamine-mercaptopurine*. A 9-month-old infant presented with nephropathic cystinosis and was prescribed mercaptopurine by the GP instead of mercaptamine. After a month on the wrong medication, she developed pancytopenia but ultimately made a full recovery (MHPRA 2010).

b. *hydromorphone-morphine*. An elderly patient was discharged after being administered hydromorphone instead of the prescribed morphine by a nurse in the Emergency Department. He suffered a fatal respiratory arrest on his way home.

LASA errors are estimated to account for around 25% of all medication errors in the US (Emmerton & Rizk 2012), and occur in all aspects of medications management – during prescribing, dispensing and administration of the medication. LASA errors thus represent a significant threat to patient safety.

The bulk of extant literature on LASA errors focuses on mitigating their occurrence (Emmerton & Rizk 2012; Ghaleb *et al.* 2010, Aronson 2009; Kovacic & Chambers 2011) and very little research has been conducted into linguistic properties of the nomenclature to elucidate properties that may prime the risk of the errors occurring. Profiling of such properties could inform the name formation process and thus

prophylactically reduce the risk to patient safety. It is also possible that elucidating external factors contributing to the likelihood of confusion error (such as high syllabic similarity) will encourage reporting of adverse drug events (ADEs) and near misses, since these may be under-reported due in part to fear of reprisal, blame and reputation damage (Aronson 2009).

More needs to be known about the formal and semantic properties of the main global medication nomenclature of International Nonproprietary Names. This study examines semantic transparency in the nomenclature and the underlying conceptual taxonomy of pharmacological relationship. In the context of this study, semantic transparency is defined as the correspondence between form and meaning within a lexical unit and the extent to which meaning motivates form and meaning is derived from form.

3 Medical taxonomies and ontologies

There are many systems of classification in medicine, such as the HUGO (HUman Genome Organisation) gene nomenclature, Medical Subject Headings (MeSH) used to index published research on Medline, and the University of Washington Digital Anatomist (UWDA) (Shapiro *et al*. 2005; Segura-Bedmar *et al*. 2008). Due to the exponential growth of published research in medicine, it is now impossible for specialists to keep abreast of developments in their field, and the need has arisen to automate recognition of key concepts in the literature (Coletti & Bleich 2001, Segura-Bedmar *et al*. 2008). The Unified Medical Language System (UMLS) is an example of an ontology by which automated software can read and assimilate information in published research (Segura-Bedmar *et al*. 2008) and encompasses various nodes, such as the UWDA for anatomy. Some systems determine nomenclature, such as the HUGO gene nomenclature, and others are used to assign conceptual relations, such as the UWDA (Shapiro *et al*. 2005). The UWDA uses various semantic links, e.g. the oesophagus is *part-of* the foregut, *continuous-with* the pharynx and stomach and *adjacent-to* the trachea, thoracic aorta and thoracic vertebral column.

The terms *classification*, *taxonomy* and *ontology* are often used interchangeably to refer to any system of categorization, but for the purposes of this study, *ontology* is taken to mean any system that categorizes concepts (Stevens *et al*. 2000) and a taxonomy should be seen as a methodology for categorization. There are several key distinctions to be made. An ontology is “the concrete form of a conceptualization of a

community's knowledge of a domain" (2000: 1), whereas a taxonomy does not necessarily include added knowledge beyond the necessary and sufficient criteria for categorization. Ontologies may be multidirectional and include multiple types of semantic relation, such as meronymy and metonymy, whereas a taxonomy is an upside down tree structure (Shapiro *et al*. 2005) and is based upon intrinsic properties of its members. Taxonomies are typically 'tree-like' hierarchies, employing hyponymy (*is-a*, class membership) to express semantic relationship. In terms of Jackendoff's widely adopted theory, the organization of systems will inevitably depend upon our conceptualization of the world (Jackendoff 1983), but further consideration of that is beyond the scope of this chapter. The prototypical taxonomy is the plant or animal kingdom used in biology (Shapiro *et al*. 2005, Coletti & Bleich 2001).

According to the WHO, the INN system is a 'classification', but can be more specifically defined as a taxonomy since it only employs *is-a*, hyponymic semantic relations. Although there is a global taxonomic system for pharmaceutical substances, the Anatomical Therapeutic Chemical (ATC) index, INNs use a different taxonomy that does not align with the ATC (Segura-Bedmar *et al*. 2008) and is not used by any other organization. For example, the medication name *selegiline* in the ATC system would be found by drilling down into the taxonomy: Nervous system > Anti-parkinson drugs > Dopaminergic agents > Monoamine oxidase B inhibitors, but in the INN system by Psychopharmacologics > Antidepressants > Monoamine oxidase inhibitors.

The INN system employs at most a four-level taxonomy and assigns alphanumeric codes to each level. Although there is room for four levels, currently names fill only two levels, so the INN system can be seen as a flat taxonomy or a collection of individual taxa under an undefined hyperonym. There is sparse information on the taxonomy beyond the statutory guidance of the WHO and neither definitions nor necessary and sufficient criteria for inclusion in the taxonomy are provided. The INN system is unique in the world of medical ontologies and taxonomies in that the nomenclature it motivates is used by people at all levels of society who hold varying levels of knowledge.

4 A typology of taxa in the INN nomenclature

Pharmacological relationships between substances are demonstrated by the use of a common stem (WHO 1997: 1), which may be a prefix, infix, suffix, or a 'freefix'. By using a common stem, the INN indicates that its denoted substance belongs to a group of substances with similar

pharmacological activity (WHO 1997: 1). The common stem or sub-stem is combined with a "random, fantasy prefix", normally chosen by the submitter of the new substance (WHO 1997: 6) and "the only requirement is to contribute to a euphonious and distinctive name" (WHO 2004: 128). Displaying taxonomy from right to left, starting at the end of the name, is a predictable approach for the user as they can first categorize the name under its stem and further sub-categorize under sub-stems by reading to the left. The reverse would be impossible due to the meaningless prefix. The INN taxonomy is based upon hyponymy, and in this chapter, *stem* will be used to denote hyperonym and *sub-stem* to denote hyponym.

This chapter presents a qualitative typology of taxa found in the INN nomenclature and reviews the implications of these types in the usability of INNs. WHO guidance stipulates that names must not be liable to confusion and that relationship must be shown by the use of a common stem. Therefore, there must be a robust and structured underlying conceptual taxonomy in place to facilitate correct usage of the medication names. The typology that follows is a qualitative analysis of the author's database of monolexic INNs (n=7,111) and the *WHO Stem Book* 2011, which provides information on the INN taxonomy and lists of INNs containing each stem and sub-stem (WHO 2011).

4.1 Single-level taxa

There are many INNs that are regularly formed, some with only a single-level taxon represented by a single stem. This type of taxon has no hyponyms. Examples are given in Table 1-1 overleaf. These stems occur as all four types of affix: prefix, infix, suffix and freefix.

These single-level taxa illustrate the longevity of the INN nomenclature: from its inception in 1952, the taxonomy has allowed for developments in pharmacology by creating empty pharmacological taxa. Stems are created, but may not appear in names immediately – the system is proactive rather than reactive. This future-proofing is similar to Dmitri Mendeleev's periodic table, in which gaps were left for elements not yet discovered. It is possible that in future a sub-category of cannabinoid receptor agonists may be discovered and in that case a sub-stem of *nab* can be created.

Stem	*Affix type*	*Pharmacology*	*Examples of INNs*
arte-	prefix	antimalarial agents, artemisinin related compounds	arteflene, arterolane
-coxib	suffix	selective cyclo-oxygenase inhibitors	etoricoxib, tilmacoxib
-formin	suffix	antihyperglycaemics, phenformin derivatives	benfosformin, metformin
nab	freefix	cannabinoid receptor agonists	menabitan, nonabine
-pris-	infix	steroidal compounds acting on progesterone receptors	ulipristal, asoprisnil

Table 1-1: Examples of single-level taxa[2]

4.2 Regular taxa

Many stem taxa clearly display their taxonomy in names that can be interpreted from right to left. The stem is the suffix and sub-stems are distinguished from their co-hyponyms as infixes directly before the suffix stem. The taxon for "antiasthmatic, antiallergic substances not acting primarily as antihistaminics" has the stem *-ast*, and sub-stems *-lukast*, *-milast*, *-trodast* and *-zolast*. Montelukast is a substance in this group and its meaning can be easily derived from the order of word parts: the suffix stem *-ast* can be used to categorize the substance as part of the antiasthmatic taxon and the infix *-luk-* can be used to further sub-categorize it as a leukotriene receptor antagonist.

In regular taxa such as these, morphemic concatenation is ordered as in Table 1-2.

[2] The hyphen indicates the position of the affix in the name. Freefixes are unhyphenated to indicate they can appear in any position.