Analysis and Synthesis

Interdependent Operations in Chemical Language and Practice

Claus Jacob

Abstract: Chemical symbolism provides the linguistic representations for experimental research. It is based on an empirical set of formal (syntactic) rules that allows operations on formulas and reaction equations. The semantic interpretation of formulas and reaction equations links these operations to experimental analysis and synthesis. These syntactic and semantic aspects of chemical symbolism guide as well as limit chemical research. A better understanding of these aspects of chemical language allows chemists to rationalize novel approaches to chemical research (*e.g.* combinatorial chemistry) and possibly exploit the vast area of 'surprise discoveries'.

Keywords: *chemical language, syntax, synthesis, analysis, combinatorial chemistry.*

1. Introduction

Chemistry is an experimental science that transforms both substances and (chemical) language. On the one hand, chemists analyze and synthesize new compounds in the laboratory; on the other, they make analytical and synthetic statements about these compounds in research articles. Consequently, language is an essential aspect of chemistry and there can be no doubt that chemical language in more than one way influences the course taken by chemical research.[1] It is therefore essential to understand how chemists' use their language, what rules govern its use, and what consequences the utilization of this language has for chemistry as a whole. While it is essential to distinguish between chemical experiments and chemical language, it is equally important to distinguish between different 'levels' of chemical language. Chemistry employs a particular language to *name* its research objects ('substances'). It provides a vocabulary to talk *about* substances. Additionally, it is possible to discuss substances in general terms, converse about laws, models, and theories that govern the behavior of elements and compounds. On yet

another level, it is then possible to enter an epistemological discussion about theories, their origin, and empirical basis.

All levels of chemical language are vital for chemical research. In particular, the relationship between the chemical symbols used to represent substances and the substances themselves is one of the most central for the research chemist. It is at this interface between the manipulation of substances and the manipulation of symbols that simple operations (such as mixing, burning) become describable and generally reproducible, become part of a science.

The next section therefore takes a closer look at the different levels of chemical language. Section 3 defines chemical symbolism as a language. Section 4 investigates the empirical basis of chemical symbolism. Section 5 studies the interdependence between the different operations of analysis and synthesis on the bench and on the blackboard. Section 6 discusses the influence language has on the progress of chemical research in general and the potential limitations the use of a specific chemical language poses for research in particular. Section 7 explores possibilities of a more dynamic relationship between chemical practice and language (*e.g.* combinatorial chemistry, computer simulations). The eighth and last section briefly reviews the strengths and weaknesses of present-day chemical language and considers the future of 'random' experimentation.

2. Different levels of chemical language

While chemical symbolism has attracted considerable attention, it is important to keep in mind that chemical language consists of several different levels that exhibit increasing degrees of abstraction from chemical practice. Most discussions of chemical language focus on one of these levels. In order to understand the specific rules that apply to each level it is helpful to discuss these languages separately. The different 'sub-languages' used on these levels exhibit distinctive linguistic and epistemological properties and should not be confused with each other. Nevertheless, all of the languages used in chemistry can be studied as *languages*.

The initial level of chemical language contains chemical symbols for substances and, at first glance, hardly resembles a modern language. *Chemical symbolism*[2] has its very own rules regarding the operational use of symbols. Treatment of symbolism as language implies, for example, that it is possible to define formal and semantic rules for the use of chemical symbols. This initial choice – although not entirely unproblematic – permits the application of basic linguistic concepts to all levels of chemical language. It also allows a de-

tailed analysis of the strengths, pitfalls, limitations, and potential areas of improvement of chemical symbolism.

Similarities to ordinary languages are more apparent on other levels that are equally important in chemistry. The second level provides a vocabulary that enables chemists to talk about substances in general. With respect to symbolism, it represents a kind of 'meta-language' ('language of chemical abstractors' in the protochemical sense). It contains 'ideators' and 'abstractors' such as »element« and »compound« (Psarros 1996). For example, the statements »sodium and potassium are elements« or »the compound is 98% pure (by GC analysis)« are part of this particular language. Therefore, chemists are able talk about a vast number of substances simultaneously. Instead of naming 110 individual elements, »element« can be used to talk about all elements at the same time. The definition of these terms is of utmost importance. Chemists continuously introduce new terms (*e.g.* »fullerenes«, »Lewis acids«) that require universally accepted definitions.

The language of this level is a modified (*i.e.* specialized) ordinary language, *e.g.* English or German. Semantic rules are of utmost importance for the precise (and specialized) definition and successful use of terms. 'Protochemistry' has attempted an operational definition of most of the more fundamental 'abstractors' like »element« and »compound« (Janich 1994, 1996; Psarros 1995; 1999, pp. 68-129; Hartmann 1996). This level of language is essential for effective chemical communication. In addition, the use of these terms is a pre-condition for the formation of general chemical theories.

The third level of chemical language contains terms that are used to discuss 'abstractors' such as »element« and »compound« (as defined on the second level) as part of laws, models, and theories in a general context. For example, the laws of constant and multiple proportions are part of this 'language of chemical theory'. This language is similar to the one used on the second level and is a modified (*i.e.* specialized) ordinary language. Unlike the language utilized on the second level, its use is not limited to chemistry. Theories and laws are used in most sciences and can be discussed in a more general context.[3]

On yet another level, it is possible to enter an epistemological discussion about chemistry as a whole (*e.g.* about chemical theories, their origin and empirical basis). Statements such as »a reaction mechanism is the linguistic representation of a controlled (chemical) reaction« belong to that level and this article itself is written in this language. The language at the fourth level is the language of philosophy. It includes particular syntactic and semantic problems that, although highly important, cannot be discussed here.

It is obvious that the whole situation is even more complex. For example, each 'sub-language' is connected with the other ones and the language on one level defines entities for the language on the next level. For example, »Na« (L1) belongs to the elements, »element« (L2) is discussed in chemical hy-

potheses, »hypotheses« (L3) can be tested and falsified in the Popperian sense, and thus enter L4. This raises interesting semantic problems for each level of chemical language. A detailed discussion of these levels and their particular degrees of abstraction will therefore be important in the future, especially since some chemists tend to treat their language as one entity. Nevertheless, it is sufficient for this particular discussion to distinguish between chemical symbolism and the other levels. The level of chemical symbols, formulas, and reaction equations is perhaps the most interesting one from the chemist's point of view. A particular chemical experiment (*i.e.* an analysis or synthesis) is 'represented' by chemical symbols and reaction equations, not by words such as 'element' or 'chemically pure'. Therefore, chemical symbolism demands a more detailed discussion as an important chemical 'sub-language'.

3. Chemical symbolism as language

The particular kind of chemical language that is used to denote compounds, their properties, and conversions is at the heart of chemistry. Interestingly, a number of philosophical reflections upon chemical language (*e.g.* 'Proto-chemistry') have focussed on the more abstract second and third levels. Indeed, it is not self-evident why chemical symbolism should (or could) be treated as a language at all. After all, there are no words or sentences in the classical sense. Nevertheless, recent attempts to approach chemical language from different perspectives have also included discussions about the formal properties of chemical symbolism, *e.g.* in the wider context of semiotics (Schummer 1996).

In order to discuss chemical symbolism in more detail, its specific linguistic characteristics have to be defined. Basically, it consists of an 'alphabet', a particular syntax, and a set of semantic rules. This initial definition is important for the following analysis. Although that is not necessarily the only possible approach towards chemical symbolism, it can still be used to discuss important aspects of chemical research in general and chemical communication in particular.

The individual linguistic elements of chemical symbolism can be defined in analogy to a 'model language' consisting of an alphabet of elemental symbols that all carry a particular meaning. The elemental symbols are then connected to form 'words', according to orthographic rules; and words are connected to form 'sentences', according to grammatical rules. Both formal rules are summarized as syntactic rules to distinguish them from semantic rules which govern the meaning of elemental symbols, words, and sentences. Such a 'model language' is different from ordinary English or German.

The presently used *chemical alphabet* consists of approximately 110 *symbols* representing the known chemical elements (from »H« to »Uno«). However, the number of these 'elementary' symbols is not limited since new symbols can be introduced. Elemental symbols can be combined in order to form a chemical formula (*e.g.* »NaCl«) and reaction equations (*e.g.* »2 Na + Cl$_2$ → 2 NaCl«). These combinations of symbols follow a set of *formal rules*, comparable to the rules that govern the formation of words and sentences of a 'model language', and will be defined as *chemical syntax*.[4] Chemical syntax covers empirical rules regarding 'valency', 'oxidation state', 'electronegativity', 'affinity' and 'reaction mechanisms' that have found their way into chemical theory (Psarros 1996). It is possible to distinguish between a *chemical orthography* and a *chemical grammar*. Chemical orthography provides the rules that govern the combination of elemental symbols to chemical formulas (*e.g.* valency, oxidation state). It determines which elemental symbols can be combined, in which ratio and how. For example, the symbols »Na« and »Cl« can be combined to »NaCl« using the rule that 1 »Na« can be combined with 1 »Cl« according to rules about valency and oxidation state of the denoted elements. Chemical grammar provides the rules that govern reaction equations. It determines the stoichiometric coefficients ('balanced' equations), the use of a unidirectional or an equilibrium arrow, and 'reaction conditions' as long as they are part of a reaction equation (*e.g.* solvent, temperature). Chemical orthography and grammar are closely related. For example, the grammar rules of the reaction formula »2 Na + Cl$_2$ → 2 NaCl« are determined by the orthography of»Na«, »Cl$_2$«, and »NaCl«. It is therefore less confusing to call all formal rules chemical syntax and avoid expressions such as orthography and grammar at this stage of the discussion.

It is, however, necessary to define one further aspect of chemical symbolism. Each symbol, formula, and reaction equation has a 'meaning' in the world of substances. The relationship between »NaCl« and a lump of salt is a semantic problem at the heart of many epistemological discussions. *Chemical semantics* discusses the 'meaning' of linguistic representations (*e.g.* symbols, formulas, or reaction equations) in relation to chemical practice.

Chemical semantics is important and most recent discussions have focussed on this aspect of chemical symbolism. While chemical semantics is ideally suited to describe the relationship between existing substances and their linguistic representations, chemical syntax enables chemists to form new symbols as representations of substances not yet synthesized. The present approach intentionally separates the syntactic from the semantic aspects of chemical symbolism. The meaning of »NaCl«, *i.e.* common salt with all its (chemical, physical, social, and cultural) properties, is independent from both the 'orthographic' correctness of »NaCl« versus »Na$_3$Cl« and the grammatical correctness of »2 Na + Cl$_2$ → 2 NaCl« versus »Na + Cl$_2$ → NaCl«. *More*

importantly, the syntactic correctness of a formula is independent of its meaning,[5] for, once syntactical rules have been established, we can correctly create new formulas without bothering about their meaning. The clear distinction between syntactic and semantic rules allows for an important *asymmetry* between operations with language and operations with compounds.

This asymmetry is the basis of planning new reactions in chemistry and it is an important aspect of chemical research. It allows for a wealth of chemical formulas and reaction equations to be proposed even before a single test-tube has been filled. It delivers *a priori* formulas and reaction equations that are solely generated by syntactic rules while, at this stage, ignoring the empirical implications (*i.e.* 'meaning') of these formulas and equations. For example, writing a new formula »NaAt« is based on syntactic rules (valency, analogy with »NaCl« or »NaBr«). Although »NaAt« does not represent a compound that can be made in practice, it does not violate syntactic rules and – as a chemical formula – is available for further chemical research.

In summary, the distinction between syntactic and semantic properties of chemical symbolism allows the introduction of chemical formulas that are syntactically correct but do not (yet) have an empirical basis. Such formulas would be forbidden in a language governed by the semantic requirements that each formula most represent something.

It is now possible to further distinguish between operations with substances and operations with symbols. Among operations with chemical symbols, we can define *analysis$_1$* and *synthesis$_1$* (Jacob 1998, p. 38-40). Analysis$_1$ and synthesis$_1$ are guided by formal (syntactic) rules, once established on an empirical basis and now part of chemical theory (*e.g.* rules of valency, oxidation states, functional groups, reaction mechanisms, *etc.*). Their outcome is a (linguistic) proposition in form of an *analytical$_1$* or *synthetic$_1$ statement.*[6] An analytical$_1$ statement can be made without prior empirical research. It extracts information that is already present in the original formula or reaction equation. For instance, the statement »2 moles of NaAt contain 2 moles of Na and one mole of At$_2$« (*i.e.* »2 NaAt → 2 Na + At$_2$«) is an analytical$_1$ statement exclusively derived from syntactically correct operations on the formula »NaAt«. Such statement does not require the immediate chemical analysis of the denoted compound. The statement »francium and astatine form francium astatide« (*i.e.* »2 Fr + At$_2$ → 2 FrAt«) is a synthetic$_1$ statement. Analytical$_1$ and synthetic$_1$ statements equally apply syntactic rules. Analytical$_1$ statements frequently predict the chemical properties (*i.e.* composition) of a compound based on the symbols present in its linguistic representation. Synthetic$_1$ statements, on the other hand, often combine symbols to form new formulas that 'represent' yet unknown compounds.

On the other hand, the practical chemical operations of analysis and synthesis (as defined in chemistry) generate *chemical compounds*. These experi-

mental operations are performed on substances, not on symbols. They are guided by experimental rules describing chemical practice (*e.g.* the correct use of chemical equipment, purification methods). Experimental operations involve controlled mixing of well-defined starting materials, purification methods (*e.g.* chromatography, recrystallization) and analytical methods (*e.g.* mass spectrometry, elementary analysis). These operations will be defined as *analysis$_2$* and *synthesis$_2$*. Their actual outcome is an *analytical$_2$* or *synthetic$_2$* *compound* (Jacob 1998, p. 38-40).[7] Analysis$_2$ and synthesis$_2$ are part of chemical research and their planning and execution depends on a framework of chemical theory. Although similar operations are possible outside chemistry (*e.g.* cooking, mixing of mortar), the latter will not be defined as analysis$_2$ and synthesis$_2$ but as 'random mixing' (see Sect. 7).

These two sets of operations and their actual outcomes are not simply related by the coincidental use of the same terminology. The use of specific elementary symbols for chemical elements *intentionally* allows the linguistic operations of analysis$_1$ and synthesis$_1$ to be carried out – while chemical practice performs the practical operations of analysis$_2$ and synthesis$_2$. Therefore, chemical language heavily depends on its empirical basis.

4. The empirical basis of chemical syntax

Ordinary languages use symbols and syntactic rules based on convention and more or less rational design. It is also possible to invent new languages based on new 'universal grammars' (*e.g.* Lightfoot 1999, pp. 49-76). In contrast, chemical symbols, formulas, and equations *as well as* the syntactic rules of chemical symbolism are mainly based on experimental chemical experience.

Superficially the combination of »Na« and »Cl« to »NaCl« is similar to the combination of »screw« and »driver« in »screwdriver«. While the latter example, however, simply represents a way to describe a new instrument by the combination of the names of two known entities, the combination of »Na« and »Cl« to «NaCl« not only names a particular compound, it also tells the chemist its (empirical) composition and how to make it. Take one part of Na and one part of Cl and the result being one part of NaCl! No such predictions or construction instructions can be made in the screwdriver example.

Therefore, a chemical formula is related to a chemical compound in *two ways*. First, the formula represents a compound ('linguistic representation'). This relationship is governed by semantic rules and is frequently at the center of epistemological discussions. For example, »NaCl« represents purified rock salt. Second, the chemical syntax is based on rules that have their (empirical) origin in the world of compounds and chemical reactions. For example, elementary analyses of many different salts have determined the rule 'Na is

monovalent with oxidation state $+1$ in sodium halide compounds'. This second 'link' between a compound or reaction and their linguistic representations (formula, reaction equation) enables chemists to claim experimental relevance for the outcome of operations with symbols.

A closer look at the NaCl example will shed some light on this connection between analysis$_1$ and analysis$_2$: The statement »NaCl contains one (atom) equivalent of sodium and one (atom) equivalent of chlorine« is an analytical$_1$ statement. At the same time, it is also an empirical statement derived from and supported by the finding that electrochemical decomposition of melted sodium chloride (analysis$_2$) yields one equivalent of sodium and one equivalent of chlorine. While the original statement is made on the basis of practical experience (analysis$_2$) further chemical predictions concerning NaCl can be made by analytical$_1$ or synthetic$_1$ statements simply based on the representation »NaCl«. No additional laboratory work is required to make those predictions – although the testing of those predictions might require further experiments. For example, the chemist can confidently enter the laboratory with the clear expectation that his/her melted salt will yield sodium and chlorine, but not potassium or bromine. He/she also 'knows' that the compound named by »NaCl« consists of sodium and chloride ions and will react with silver nitrate to form a white precipitate of silver chloride. On the other hand, the chemist may employ the syntactic rules of chemical language that govern chemical reaction equations and state a synthetic$_1$ route to »NaAt« or predict a precipitate when NaAt reacts with silver nitrate long before NaAt has ever been synthesized$_2$ in a laboratory (»NaAt + AgNO$_3$ → {AgAt}$\downarrow$ + NaNO$_3$«). Although such *a priori* reaction equations are possible, it is not possible to know if they are meaningful in practice. Analysis$_1$ and synthesis$_1$ cannot replace analysis$_2$ and synthesis$_2$.[8]

Differences in formal and semantic aspects of ordinary and chemical language also explain the difference between the word »screwdriver« and the formula »NaCl«. Analysis$_1$ or synthesis$_1$ of »screwdriver« will not yield any information about the components (analysis$_2$) or manufacture (synthesis$_2$) of this tool (but – *coincidentally* – about its use). The formal (orthographic) rules that govern the separation of »screwdriver« into »screw« and »driver« are independent of the empirical separation of a screwdriver into a rod of iron and junks of plastic. In addition, the meaning of »screwdriver« is not linked to the meaning of »screw« or »driver«. There is no apparent relationship between the two types of analysis or synthesis in this common-language example.

It is important to mention that there are different kinds of chemical formulas that contain different, and a different degree of, information and hence lead to different analytical$_1$ results. The simple empirical formula »C$_2$H$_2$Cl$_2$« can be rewritten as »CHClCHCl« or »CH$_2$CCl$_2$«. By rearranging the element symbols according to certain rules, it is possible to obtain more detailed

analytical[1] knowledge about that particular compound. A further refinement would be the use of stereostructures that indicate if the compound is either *cis* or *trans*.[9] Additionally it is possible to specify which isotopes of carbon, hydrogen, and chlorine are present, what the polarity of the bonds is, and how stable the molecule is. The 'refinement' of empirical structures has been one of the main tasks of research chemists since the introduction of modern chemical symbolism at the end of the 18th Century (Crosland 1962, pp. 177-193; Hudson 1992, pp. 69-70; Bensaude-Vincent & Stengers 1996, pp. 87-91). It involves the refinement of syntactic rules (based on experimental experience) and the introduction of general chemical laws and theories (*e.g.* reaction mechanisms).

It is not the aim of this publication to redraw or improve chemical symbolism. However, it is crucial to understand that the use of a particular formula allows only a particular analysis[1] using a particular set of syntactic rules based on certain chemical laws. If an empirical formula like »$C_2H_2Cl_2$« is used, analysis[1] cannot lead to statements about functional groups, polarity, isotope ratio or stereochemistry. If »CHClCHCl« is used, there is still no analytical[1] information about *cis* or *trans* configuration and hence the dipole moment. Of course, this information is likely to be available to modern-day chemists, *but it is not always present in the type of representations used*. It is therefore one thing to point towards the overall knowledge generally available to chemists and another one to look at the precise analytical[1] statements that can be derived from a formula actually used in that particular instance.

Chemists use analytical[1] and synthetic[1] formulas and reaction equations to predict analytical[2] fragments, novel synthetic[2] compounds and the direction of chemical reactions in the laboratory. This language is very powerful since it allows chemists to derive statements about 'compounds' that have actually never been produced in a laboratory. For example, »H« and »O« can be combined in numerous ways as »H_2O«, »H_2O_2«, »HO_2«, *etc.* Coincidentally, these compounds have also been made in the laboratory. The combination »H_2O_{10}«, however, can be synthesized[1], but the compound H_2O_{10} (a polyoxide) has not yet been synthesized[2]. *In this respect the capacity of the language exceeds the experimental abilities of the chemist.* This might be one of the reasons why the language of chemistry has frequently been at the center of the philosophy of chemistry.

Unfortunately, the fascination with the sheer power of this language has prevented a closer look at its potential pitfalls – *especially at the potential cases where the experimental abilities of chemists might exceed the capacity of the language*. For example, there is presently no rule that would allow for the combination »H_4O_2« in synthesis[1] while such (or similar) corresponding compounds are encountered under extreme experimental conditions in gas reactions.[10] An empirically based syntactic rule (here, valency rules) can fail. It might exclude predictions of reaction products containing atoms with yet

unknown valences that could actually be synthesized$_2$ in practice. Thus, the chemical syntax allows the prediction of some new compounds and hinders the prediction of others at the same time.

This is a provocative statement that demands a detailed examination of the relationship between operations with compounds and operations with language. First, how exactly are analysis$_1$, analysis$_2$, synthesis$_1$, and synthesis$_2$ related with each other; *e.g.* are there symmetrical or asymmetrical relations? Second, do some aspects of these interdependencies hinder the scientific progress? Third, are experimental results possible that cannot be expressed in present-day chemical language – and what epistemological status or value would they have? Fourth, are these experimental results as interesting for the chemist as they might be for the philosopher of chemistry?

5. Interdependencies between analysis$_1$, analysis$_2$, synthesis$_1$, and synthesis$_2$

Analysis$_1$ and synthesis$_1$ are operations performed on linguistic representations and lead to formulas, reaction equations, and statements. Analysis$_2$ and synthesis$_2$ are operations performed on compounds and lead to other compounds.[11]

Now, consider the relation between analysis$_1$ and analysis$_2$. As already mentioned, the process and outcome of analysis$_2$ provides the empirical basis required for inventions of chemical formulas, equations and statements. It is therefore a necessary condition for analysis$_1$ that its rules are based on the practical findings of analysis$_2$. Analyses$_2$ delivers the number and associated properties of element symbols as well as the syntax to combine these symbols in an orderly fashion (*e.g.* valency, oxidation state). However, analytical$_2$ or synthetic$_2$ operations in the laboratory require a theoretical framework in order to be rationally designed and executed. Simply adding lemon juice to a fish does not explain why this operation suppresses the bad smell. If such an operation is not based on a chemical theory it could not help to test, explain, or predict observations or further operations nor might it be reproducible. Such operations are usually part of a craft (like cooking) and based on previous experience, but not considered scientific. In particular in chemistry, such practical experience soon reaches its limits because it does not allow chemists to make predictions about the outcome of an unknown reaction or the possible synthetic$_2$ route to a new compound (*e.g.* retro-synthesis).[12]

Therefore, the practice of analysis$_2$ is driven – and potentially limited by – the outcome of a prior analyses$_1$ based on chemical formulas (*e.g.* the search for H_2 and C during thermal decomposition of CH_4, the fragment CH_3 in

the mass spectrum of methanol). The predictions made by analysis$_1$ are, however, neither sufficient nor necessary presuppositions of analysis$_2$: not sufficient, because analysis$_2$ frequently brings to light unexpected compounds or 'impurities'; not necessary, because analysis$_2$ might reveal a set of products completely different from those expected for a given compound. For example, the analysis$_1$ of the empirical formula »C_2H_6O« might predict the presence of an ethyl and a hydroxyl group (alcohol). However, analysis$_2$ could show that there are other reactive groups in the sample (ether) or that there might not be any hydroxyl groups present at all. This indicates that neither the chemical language nor the chemical practice is independent from each other. Simply shifting the focus to only one aspect – practice or language – hides the interdependence of both.

A similar interdependence can be found in the case of synthesis$_1$ and synthesis$_2$. Chemical formulas allow the invention of chemical reaction equations that make predictions about the formation of new compounds. These equations can then be used in practice where they might stimulate the synthesis$_2$ of a new compound. There is no guarantee, however, that this synthesis$_2$ will indeed deliver the expected compound. Predicting a compound by synthesis$_1$ is not sufficient to guarantee its synthesis$_2$. Numerous well-designed reactions have 'gone wrong' in the past. Prior synthesis$_1$ is – as so-called 'surprise discoveries' sometimes show – not even a necessary presupposition of synthesis$_2$.

This is a central aspect of the relationship between linguistic representations and represented compounds. Representations obtained by analysis$_1$/synthesis$_1$ are neither sufficient *nor* necessary presuppositions of the outcome of analysis$_2$ or synthesis$_2$. They are justified by syntactic rules but do not necessarily have referents among actual compounds. Operations with such representations provide a useful tool for research chemists, but their outcome is not absolutely reliable as the outcome of most analyses$_2$/syntheses$_2$ indicates (unsuccessful attempts, by-products, and 'accidental surprise discoveries'). This aspect will be discussed further in the next section where possible scientific limitations caused by chemical language are discussed.

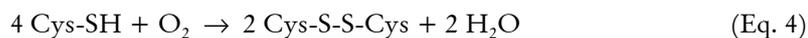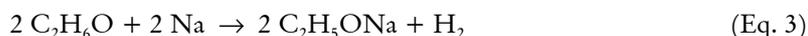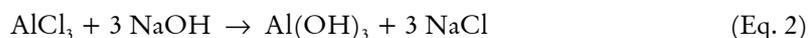# 6. Scientific limitations caused by analysis$_1$ and synthesis$_1$

An apparent restriction of the chemical language is the limited number of entries in its 'alphabet'. There are at present about 110 symbols for chemical elements. Any predictions made by chemical statements will be limited to

these known elements. Thus, it is *a priori* impossible to design a *chemical* reaction equation (synthesis₁) that would lead to a new element as one of its products. The *deliberate discovery of a new element* is excluded.

On the other hand, the use of a symbol (or name) for an element that does not exist (historical examples are »phlogiston« and »muriatium«) represents another aspect of chemical language that might lead to flawed experimental results. In this case, analysis₂ or synthesis₂ guided by analysis₁ or synthesis₁, respectively, would lead to inconclusive, 'surprising', or dubious results where elements turn out to be compounds and *vice versa*. In the late 18ᵗʰ century, the limitations of chemical research due to the lack of names like »oxygen« and the presence of names like »phlogiston« caused misinterpretations of experiments and led to fruitless attempts to isolate phlogiston and hence partially hindered scientific progress for a number of decades.

Present day chemistry, however, hardly experiences problems based on an incomplete or faulty chemical alphabet. This is due to the systematic arrangement of elements in the periodic table – based on the physical properties of their atoms. The periodic system and the relationship between an element and its number of protons allow the discovery of new elements and might even predict some of their properties. They also rule out new additional elements with proton numbers of less than 110 (Shriver *et al.* 1998, pp. 3-49).

While the chemical alphabet is easily corrected, the limitations caused by chemical syntax are not. The following simple examples further illustrate chemical syntax and syntax problems. Some of these difficulties will appear trivial to chemists because they can simply be resolved by using another type of formula (see Sect. 4). However, these examples shall illustrate general difficulties associated with analysis₁ of various types of formulas.

$$H_2 + O_2 \rightarrow H_2O_2 \qquad\qquad\qquad\qquad\qquad (Eq.\ 1)$$

$$AlCl_3 + 3\ NaOH \rightarrow Al(OH)_3 + 3\ NaCl \qquad\qquad (Eq.\ 2)$$

$$2\ C_2H_6O + 2\ Na \rightarrow 2\ C_2H_5ONa + H_2 \qquad\qquad (Eq.\ 3)$$

$$4\ Cys\text{-}SH + O_2 \rightarrow 2\ Cys\text{-}S\text{-}S\text{-}Cys + 2\ H_2O \qquad (Eq.\ 4)$$

$$60\ C \rightarrow C_{60} \qquad\qquad\qquad\qquad\qquad\qquad (Eq.\ 5)$$

Equation 1 is a synthetic₁ statement describing the synthesis₂ of H_2O_2. As it stands, however, this equation is rather useless in chemical practice. Direct oxidation of hydrogen under oxygen yields water, but not hydrogen peroxide. The equation follows the syntactic rules – but it cannot be transformed into a successful experiment. Synthesis₁ simply predicts the *wrong synthetic₂ product*.

Equation 2 is related to a similar problem. Although aluminium hydroxide is a possible product of this reaction, the outcome critically depends on the precise reaction conditions used. The reaction of $AlCl_3$ with NaOH can lead to the formation of $Na[Al(OH)_4]$ as well as $Al(OH)_3$. Equation 2 only predicts one *possible product* of the reaction of $AlCl_3$ with NaOH.'.

Equation 3 contains expressions for isomeric compounds (ethanol or dimethylether). Such empirical formulas ('C_2H_6O') are equivocal. Therefore Equation 3 might be valid for synthesis$_2$ (in the case of ethanol) – or not (in the case of dimethylether). The formulas used in Equation 3 do not contain sufficient analytical$_1$ information to name one particular compound and therefore invalidate the predictions. Equation 3 predicts one *possible product*.

Equation 4 (an example from biochemistry) represents a similar problem. First, the oxidation of Cys-SH (cysteine) might not lead to the disulfide but to a sulfenic, sulfinic, or sulfonic acid (Jacob *et al.* 1998, 1999). Equation 4 therefore might predict the wrong product. Secondly, Equation 4 does not specify if L-cysteine or D-cysteine is oxidized. Again, the experimental outcome might be different for both isomers (for example, if the reaction is enzyme catalyzed). Equation 4 does not make a compelling prediction for a practical experiment. 'Surprising' results are possible if the wrong isomer is used – or a new type of isomerism is present that has not yet been discovered. This means that such synthetic$_1$ equations can be used to shed light only on a *particular aspect* of a chemical reaction. If D,L-isomerism is not an issue in synthesis$_1$ it also remains unspecified in synthesis$_2$. Increasing the range of aspects addressed in synthesis$_1$ (structural formulas, isomerism, and energy and stability considerations) is one of the main occupations of research chemists.

It is the common feature of these examples that the formulas and syntactic rules used are not sufficient to predict the precise outcome of an experiment. There is always some experimental information missing when simply looking at reaction equations, formulas of compounds, or syntactic rules. It is therefore not entirely correct to treat a chemical reaction equation simply as a calculus (Psarros 1996). This would only reflect synthesis$_1$ but completely deny the practical aspects involved in synthesis$_2$. Even if a calculus as part of synthesis$_1$ delivers valid chemical statements, that validity is a logical and not a chemical (practical) one. It has long been known that treating a language as a calculus only reflects formal aspects of that particular language. Additionally, however, the symbols have meaning and potentially "a close relation to actions and perceptions" (Carnap 1937, p. 5). This 'close relationship' is of particular importance for chemical language and leads to semantic aspects of chemical symbolism.[13] Although powerful, the chemical language is prone to make wrong predictions and is a limiting factor when it comes to the 'discovery' of new compounds.

We can understand the discovery of new products only if the synthesis$_2$ is reformulated by means of the corresponding chemical equation (synthesis$_1$). This is shown in Equation 5. The synthesis$_2$ of fullerene is a simple chemical procedure that could have been accomplished more than a hundred years ago (Kroto *et al.* 1985, Krätschmer *et al.* 1991). Why did it take until the 1980s to obtain C_{60}? The answer cannot be one concerning chemical practice alone. Rather it is related to the rules of chemical syntax too. The expression »C_{60}« has simply not been part of chemical nomenclature. There were – until the 1980s – no syntactic rules that would allow for synthesis$_1$ of »C_{60}«; the reaction equation »60 C $\rightarrow$ C_{60}« would have been syntactically wrong. The outcome of such a synthesis$_1$ would have been a representation of either graphite or diamond or of any combination of C_1, C_2, C_3, C_4, *etc.* Therefore, the term »C_{60}« did not occur in synthetic$_1$ equations and, thus, nobody attempted to synthesize$_2$ fullerenes *intentionally*. Surprisingly, however, C_{60} later proved to be thermodynamically stable *in practice*, unlike any other C_1, C_2, *etc.*

These examples show that chemical symbolism clearly promotes *and* limits chemical research. Language does not simply 'reflect' or 'record' our knowledge about substances; it also influences the direction taken by chemical research. 'No name – no game' would summarize this often forgotten relationship. Again, the crucial – but allowed – asymmetry between the operations of analysis$_1$/synthesis$_1$ on the one hand and analysis$_2$/synthesis$_2$ on the other becomes apparent when the formal rules of chemical syntax are clearly separated from the semantic meaning of a formula or reaction equation. To emphasize this point: meaning is only one link between symbolism and experiment. The empirical basis of syntactic rules is the other.

What consequences can be drawn from these interdependencies? Is there a way to improve chemistry by changing the chemical language, its nomenclature or syntax? Or is it possible to redefine the relationship between language and practice (*i.e.* semantic aspects) in order to advance chemical research?

Most of these questions have to be addressed by chemists. Some aspects, however, are of a philosophical nature and can be clarified here. It is of paramount importance for chemists and philosophers to discuss the interdependencies between chemical practice and chemical language. The clear awareness of these problems is a *conditio sine qua non* for avoiding wrong experimental predictions or unnecessary limitations based on chemical language. It has to be kept in mind that the chemical nomenclature is expandable and that the syntax that rules the design of reaction equations is imperfect.

Moreover, there is a potential pragmatic circle of chemical research to be fully understood. Experimental evidence feeds chemical language (repertoire of symbols, syntactic rules) that then predicts the outcome of further experiments. This circle has a number of implications – not all of them necessarily negative.

First, it supports the integrity of chemistry since it directs chemical research in one specific direction. Chemical research would be impossible without the interdependence of theory (language) and practice. The universal use of only *one* chemical language also creates a 'closed society' of chemical practitioners and allows for one unified world-wide chemistry. Any 'chemistry' outside the traditional language and practice will be considered as an unscientific heresy and hence suppressed (Kuhn 1996, pp. 43-51). Experiments that do not comply with the standard rules of chemistry are not regarded chemical research (*e.g.* alchemy); neither are theories that do not put their predictions to the (experimental) test (*e.g.* metaphysics).

Second, the interdependence between analysis$_1$/synthesis$_1$ and analysis$_2$/synthesis$_2$ enables the rapid expansion of chemical practice and chemical symbolism. Expansion of symbolism occurs in a controlled manner with a solid empirical basis. At the same time, it stimulates the expansion of the empirical basis itself. Hence, the interdependence is a major driving force behind present-day chemistry.

Third, the unity of chemical language does not rule out subdiscipline formation (*e.g.* biochemistry, quantum chemistry) and the defined use of specialized additional languages in those disciplines. Such additional languages (*e.g.* the language used in biochemistry to describe *in vitro* experiments, the mathematical representations used in quantum mechanics to describe wave functions) can be used to address specific aspects of a subdiscipline that cannot be described with chemical symbolism alone. In this respect, the focus on symbolism might hinder the rapid development of chemistry as a science. For example, replacing the chemical symbolism with the wave notations of quantum mechanics and stability calculations might have led to the discovery of C_{60} at a much earlier stage.

Chemical symbolism not only guarantees undisturbed research but also inhibits unconventional thinking. Any experimental design that does not fit traditional chemical rules is doomed. If a compound cannot be synthesized$_1$ on paper, why should a chemist try to carry out a synthesis$_2$ in the laboratory that follows such an apparently 'impossible' pathway? The common answer is, one should not try such a synthesis$_2$. In most cases, it will indeed be impossible to succeed. If it is successful, a 'surprise discovery' is made. Only *after* the 'surprising' reactions have taken place, there is a chance of retrospective rationalization in terms of synthesis$_1$. However, this does not comply with the step-by-step approach of synthesis$_1$ followed by synthesis$_2$. What are the chances for synthesis$_2$ of a new compound that cannot even be predicted with the analytical$_1$ and synthetic$_1$ means available? The next section will address the epistemological and practical possibilities of 'unpredictable experiments'.

7. How surprising are 'surprise discoveries'?

Chemical symbols and syntactic rules for their combination are essential for the operations of analysis$_1$ and synthesis$_1$. It remains, however, unclear if the influence these rules have on the practical operations of analysis$_2$ and synthesis$_2$ is epistemologically justified or could be changed. This question aims at two opposite scenarios. On the one hand, synthesis$_1$ might propose a compound that cannot be synthesized$_2$. This is by far the most common experience in chemical research (almost considered to be the usual case) and aims at the improvement of chemical practice (synthetic$_2$ methods). It will not be discussed further at this point. More interesting are cases where the synthesis$_2$ of a new compound happens, although not predicted by synthesis$_1$ ('surprise discovery'). How can such 'surprise discoveries' be provoked, exploited, or even planned? Astonishingly for most chemists, this question is not inherently one of chemical practice but one that is related to the interdependence between language and practice. As such, the matter of 'surprise discoveries' is primarily a theoretical and not a practical one.

It is trivially true that any theoretical concept from which we draw predictions (analysis$_1$/synthesis$_1$) allows certain instances but forbids others. This does not mean, however, that the *theoretically* forbidden instances cannot occur *in practice* – because two different types of operations (analysis$_1$/synthesis$_1$ and analysis$_2$/synthesis$_2$) are involved. Therefore, there are two options that would help chemists to improve chemistry and rationalize 'surprise discoveries'.[14]

First, the chemical syntax and with it syntheses$_1$ could be improved. The more reliable and precise the rules that govern the combination of chemical symbols, the more successful will be predictions of syntheses$_2$. This option is clearly available for Equations 1-4. If precise reaction conditions (*e.g.* stoichiometry, isomerism, and structural representations instead of empirical formulas) are given, synthesis$_1$ will yield results that are more precise. In most cases (but not in all), such a refined synthesis$_1$ will also lead to a more successful synthesis$_2$. The improvement of chemical syntax is a steady process that occupies a vast number of research chemists and has a long tradition in chemistry (Hudson 1992, pp. 104-21; Bensaude-Vincent & Stengers 1996, pp. 126-159). In particular organic chemists benefit from an improved syntax since the synthesis$_2$ of approximately 36% of organic compounds is guided by reaction mechanisms (Schummer, 1997). In its most advanced version, this type of chemistry performs synthesis$_1$ with the help of computer simulations. A better 'fit' between synthesis$_1$ and the experimentally possible mainly avoids wrong predictions. While this approach aims to improve the existing syntactic rules, it does not attempt to change the interdependencies between chemical language and practice.

Second, synthesis$_2$ and the relationship between the theoretical operations and practice could be changed. In the extreme, this could lead to a kind of 'anarchy' where the distinction between a successful synthesis$_1$ and an unsuccessful one would no longer matter for synthesis$_2$. However, it is unlikely that the resulting research would still have a theoretical basis and would still yield results that can be expressed in a chemical language. A child might well generate a number of new compounds with its starter chemistry kit – but those compounds would not be recognized or further characterized. At best, chemistry would 'degenerate' to a kind of chemical craft like cooking. This type of 'chemistry' would be based on pure experience and would have a diminished scope and efficiency. It would involve synthesis$_2$ without prior synthesis$_1$ ('random mixing').

Nevertheless, this approach can be considerably improved if further experiments are performed after 'random mixing'. A chemist could pick up the child's 'mess' and carefully analyse$_2$ it – for example by gas chromatography combined with mass spectrometry. The analytical$_2$ content of the 'mess' would then be described by chemical formulas (analysis$_1$). Finally, the chemist would reconstruct a possible reaction equation to represent the child's mixing in form of synthesis$_1$. Through that, the expert chemist would redescribe the child's 'random mixing' as a (rather complex) chemical experiment only *ex post*. That procedure in a way breaks with the conventional approach of chemistry because it makes no predictions at all.

Would this type of chemistry have any use? Certainly, it would initially generate a wealth of new compounds, all of them in a way 'surprise discoveries'. Interestingly, a variation of such an approach towards chemistry is known as *combinatorial chemistry*. 'Random mixing' is coupled with a sophisticated analysis$_2$ that can be rationalized by analysis$_1$. If compounds of 'interest' are found (for example, for medical applications), 'random mixing' is used further and coupled with a process to isolate the particular compound of interest from the reproducibly generated 'mess'. Moreover, attempts are frequently made at this point to formulate a chemical equation (synthesis$_1$) to produce more of that particular compound by controlled synthesis$_2$. This involves the use of less and purer starting materials under precisely controlled reaction conditions.

However, it is important to keep in mind that this approach of 'random mixing' is not entirely free of theoretical concepts. Reaction conditions for 'random mixing' are still planned to some extent (glassware, starting materials, solvents, pressure *etc.*) and basic synthesis$_1$ reasoning provides a frame of the expected spectrum of products. In combinatorial chemistry practice, 'random mixing' is not simply combining various compounds from the chemical catalogue by *absolute chance*. Although this might ultimately lead to interesting new compounds, it is more economical to select starting materials that are known to have certain (chemical, physical, or pharmacological) prop-

erties. For example, there is a huge activity in predicting the outcomes of combinatorial chemistry on certain 'libraries'. What is different between 'random mixing' and conventional synthesis$_2$ is the absence of specific synthetic$_1$ predictions *before* the experiment.

This approach is now possible because of novel analytical$_2$ techniques. Combinatorial chemistry, as developed during the second half of the 1990s (primarily by pharmaceutical companies), is one attempt to extent chemistry beyond planned discoveries and to exploit the potential of 'surprise discoveries'. In the end, the latter are no longer surprising but sought after.

8. Conclusion

Conventional chemistry involves both analytical$_1$/synthetic$_1$ operations on the level of language (*e.g.* symbols, reaction equations) and practical operations in the laboratory (analysis$_2$/synthesis$_2$). All of those operations in many ways influence each other and thereby enable a concept-driven manipulation of substances. The interdependencies between these operations allow chemistry as a science to proceed, but at the same time also limit progress in 'unconventional' directions. Realizing that chemical theory does not completely describe chemical practice means conceding that 'forbidden' syntheses$_2$ might indeed be possible in practice. This field of planned 'surprise discoveries' therefore represents a new area of chemical research that is to a certain extent based on a new epistemological approach. It performs experimental operations *without* a detailed synthetic$_1$ basis and then harvests the fruits of a random, if controlled, chemical reaction. Future chemistry can greatly benefit from this type of combinatorial chemistry because the new approach towards synthesis$_2$ not only represents an increase in the sophistication of chemical techniques but also a new interdependency between analysis$_1$/analysis$_2$ and synthesis$_1$/synthesis$_2$.

This paper could only briefly mention some of the arising philosophical aspects of modern-day chemistry. A discussion of the levels of language used in chemistry is necessary to provide a detailed insight into the interdependencies between the different levels. In addition, the comparison with a linguistic 'model language' provides access to an important field of chemical symbolism that has not been fully explored yet.

Acknowledgements

Notes

1   To avoid a common misunderstanding: This does not imply that language is somehow more important than experimentation.

2   Use of the *expression* 'chemical symbolism' is similar to its use in chemistry since Dalton (cf. Crosland 1962, pp. 227-281; Hudson 1992, pp. 77-91). 'Symbol' is defined in Sect. 3.

3   For a detailed discussion of the more abstract levels of language and the epistemological status of abstract entities ('theory', 'model' etc.) see also Hanekamp, 1997.

4   This definition of (chemical) syntax is not identical with the definition of an English or German syntax. It is, however, appropriate for chemical symbolism.

5   This statement is about syntactic correctness. It does neither imply that the original establishment of syntactic rules is independent of their empirical basis nor that a formula or a reaction equation has no meaning.

6   The term »statement« in chemical symbolism is used here in its widest sense and includes chemical formulas and reaction equations.

7   The expression »analytical$_2$ compound« is not commonly used in experimental chemistry; here it describes compounds that have been generated for analytical$_2$ purposes (*e.g.* fragments in mass spectrometry).

8   The importance of the chemical experiment is discussed in Schummer 1994.

9   Structural formulas and stereostructures can also be treated as 'signs' implying the use of semiotic rules ('reaction mechanisms') rather than mere linguistic rules (Schummer 1996). Semiotics provides an even wider context that includes but transcends linguistics.

10   Further refinement of syntax might at some stage lead to a rule that would allow for synthesis$_1$ of this compound.

11   This comparison does not, of course, imply that it would be possible to compare chemical compounds with chemical formulas directly.

12   The issue of accidental discoveries and experiments of unpredictable outcome is discussed in Sect. 7.

13   The termini »calculus«, »symbol« and »syntax« are defined and extensively discussed in Carnap 1937). For their use in the discussion of chemical language, see Pssaros 1996.

14   The third option, the improvement of chemical research methods, has already been mentioned before and is not of interest here.

References

Bensaude-Vincent, B. and Stengers, I.: 1996, *A History of Chemistry*, Harvard University Press, Cambridge, MA.

Carnap, R.: 1937, *The Logical Syntax of Language*, Kegan Paul, London.

Crosland, M.P.: 1962, Historical Studies in the Language of Chemistry, Heinemann, London.

Hanekamp, G.: 1997, *Protochemie – vom Stoff zur Valenz*, Königshausen & Neumann, Würzburg.

Hartmann, D.: 1996, 'Protoscience and reconstruction'. *Journal of General Philosophy of Science*, **27**, 55-69.

Hudson, J.: 1992, *The History of Chemistry*, Chapman & Hall, New York.

Jacob, C.: 1998, *Protochemie – die konstruktivistische Grundlegung der Chemie* (unpublished M.A. thesis, University of Hagen).

Jacob, C.; Maret, W. & Vallee, B.L.: 1998, 'Control of zinc transfer between thionein, metallothionein and zinc proteins', *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 3489-94.

Jacob, C.; Maret, W. & Vallee, B.L.: 1999, 'Selenium redox biochemistry of zinc-sulfur coordination sites in proteins and enzymes', *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 1910-4

Janich, P.: 1994 'Protochemie', *Journal for General Philosophy of Science*, **25**, 71-87.

Janich, P.: 1996 'Chemie ohne Subjekt? Über eine Paradigmaverschiebung in der Sprache der Chemie', in: Janich & Psarros 1996, pp. 33-43.

Janich, P. & Psarros, N. (eds.), 1996, *Die Sprache der Chemie*, Königshausen & Neumann, Würzburg.

Krätschmer, W.: 1991, 'How we came to produce C_{60}-fullerite', *Zeitschrift für Physik – Atoms, Molecules and clusters*, **19**, 405-8.

Kroto, H.W.; Heath, J.R.; O'Brien, S.C.; Curl, R.F.; Smalley, R.E.: 1985, 'C_{60} – Buckminsterfullerene' *Nature*, **318**, 162-3.

Kuhn, T.S.: 1996, *The Structure of Scientific Revolutions*, Univ. Chicago Pr., Chicago.

Lightfoot, D.: 1999, *The Development of Language*, Blackwell, Oxford.

Psarros, N.: 1995, 'The constructive approach to the philosophy of chemistry' *Epistemologia*, **18**, 27-38.

Psarros, N.: 1996, 'Die chemische Reaktionsgleichung als Kalkül', in: Janich & Psarros 1996, pp. 127-138.

Psarros, N.: 1999, *Die Chemie und ihre Methoden*, Wiley, Weinheim.

Schummer, J.: 1994, 'Die Rolle des Experiments in der Chemie', in: Janich, P. (ed.), *Philosophische Perspektiven der Chemie*, BI Wissenschaftsverlag, Mannheim.

Schummer, J.: 1996, 'Zur Semiotik der chemischen Zeichensprache: Die Repräsentation dynamischer Verhältnisse mit statischen Mitteln', in: Janich & Psarros 1996, pp. 113-126.

Schummer, J.: 1997, 'Scientometric studies on chemistry II: Aims and methods of producing new chemical substances', *Scientometrics*, **39**, 125-40.

Shriver, D.F., Atkins, P.W. and Langford, C.H.: 1998, *Inorganic Chemistry*, Oxford University Press, Oxford.

Claus Jacob:
School of Chemistry, University of Exeter, Stocker Road, Exeter EX4 4QD, U.K.; C.Jacob@ex.ac.uk