# AISC Meets Natural Typography

James H. Davenport

Department of Computer Science
University of Bath, Bath BA2 7AY
United Kingdom
`J.H.Davenport@bath.ac.uk`

**Abstract.** McDermott [12,13] introduced the concept "Artificial Intelligence meets Natural Stupidity". In this paper, we explore how Artificial Intelligence and Symbolic Computation can meet Natural Typography, and how the conventions for expressing mathematics that humans understand can cause us difficulties when designing mechanised systems.

## 1 Introduction

Notation exists to be abused[1]

> the abuses of language without which any mathematical text threatens to become pedantic and even unreadable. [3, pp. viii–ix]

but some abuse is more harmful than others, and may cause real problems in a mechanised context, or even to unwary human beings.

"Semantics", in a general context, has been defined as [14, 'semantic']

> Also, (the study or analysis of) the relationships between linguistic symbols and their meanings.

In the same vein, "Notation" has been defined as [14, 'notation' 6] (which goes on to give special meanings in mathematics, music, choreography and "in other disciplines, as chemistry, logic, chess, linguistics, etc.")

> The process or method of representing numbers, quantities, relations, etc., by a set or system of signs or symbols, for the purpose of record or analysis; (hence) any such system of signs or symbols.

Another way of looking at this paper is to ask how the semantics relate to the notation.

## 2 The Trivial Differences

While occasionally embarrassing, these are cases due to a difference in conventions and, at least in theory, could be avoided by a "sufficiently clever" context mechanism. Many of them are discussed in more detail in [8].

---

[1] On 10.6.2007, a quick use of Google demonstrated 783 uses of "abus de notation", roughly 10% of which were in english-language papers.

**Intervals.** There are two well-known notations: the "Anglo-saxon" way $(0, 1]$ and the "French" way $]0, 1]$. The semantics are clear, though: in OpenMath they would be

```
<OMA>
  <OMS name="interval_oc" cd="interval1''"/>
  <OMI>0</OMI>
  <OMI>1</OMI>
</OMA>
```

**Inverse functions.** If $f$ is a many–one function $\mathbf{C} \to \mathbf{C}$, its inverse[2], which will be denoted $g$, has two possible definitions: the one–one discontinuous one, and the one–many continuous one. It is usual in Anglo-saxon cultures to denote a[3] one–one function with a lower-case initial letter, as $g$, and the one–many one with an upper-case initial letter, as $G$. Regrettably, in France the convention is apparently reversed[4]. Here the situation is worse than in the previous example: the notations are not merely different but contradictory, and any attempt at understanding them will need to know the (linguistic, in this case) context. Attempting to understand precisely *which* one-to-one function is intended seems futile, and we also note (with regret) that there is no standard notation for distinguishing between functions which differ only in their branch cuts: the author and his colleagues have generally resorted to *ad hoc* subscripts or notation such as $\underbrace{\arctan}_{\text{Derive}}$. In terms of [13, p. 150], attempting to understand this mathematically is an example of the "unnatural language" fallacy: there is no internal way of deducing which function is meant (and indeed in *some* circumstances, *some* choices of the one-to-one function may not matter).

**Metric tensor.** It is possible to define the metric tensor for flat Minkowski space as $\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ or its negative $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$. Furthermore, one can

---

[2] We use a different letter, to avoid the usual problem of iterated functions versus inverse functions.

[3] It would be tempting, but wrong, to write "the one-one function". Since it is 'obvious' that the correct inverse of $x \mapsto x^2$ as $\mathbf{R} \to \mathbf{R}$ is the positive square root, we may be tempted to think there is an obvious inverse in other circumstances. While it is normal these days to define log to have imaginary part in $(-\pi, \pi]$, the author was initially taught to have the imaginary part in $[0, 2\pi)$. [1] changed the branch cut of arctan between printings, and systems have been known to be internally inconsistent [5].

[4] Various mathematical textbooks seem to indicate this. However [2, Arcsin] gives capitals to Arcsin, Arccos and Arctan, but not to the others. There is clearly an inconsistency here, as [2, Arctan] describes arctan as the inverse function, and makes no mention of Arctan. The other inverse functions seem to have no entries in [2].

decide that the temporal variable is the last, rather than the first, co-ordinate, giving $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$ or its negative. Again, a human being generally has little difficulty with this, but it is hard to explain exactly why.

$i$ **or** $j$ This divergence of notation between electrical engineers and the rest of the world has been discussed before [7]: we merely note that we need a 'discipline' context as well as a 'linguistic' context to resolve ambiguities.

$0 \in \mathbf{N}$? This is discussed in [8]: we say here only that knowledge of linguistic context *may* help decide this question, but is far from certain.

## 3   Deep Semantics

Attempts to formalise mathematics often say "this has the usual mathematical meaning", or words to that effect. Let us look at `union` in this respect. There are three possible mathematical expressions, for which we give the MathML(-Content) and the OpenMath.

1. *n*-ary infix operator.
   **Mathematics** $a_1 \cup a_2 \cup a_3$
   **LaTeX** `a_1 \cup a_2 \cup a_3`
   **OpenMath** `<OMS name="union" cd="set1"/>`
   **MathML** `<apply> <union/> <i>`$a_1$`</i>...</apply>`
2. Acting on a set of arguments[5].
   **Mathematics** $\bigcup \{a_1, a_2, a_3\}$
   **LaTeX** `\bigcup \{a_1,a_2,a_3\}`
   **OpenMath** `<OMS name="big_union" cd="set3"/>`
   **or** `<OMS name="apply_to_list" cd="fns2"/>`
   **MathML** `<apply> <union/> <bvar>i</bvar> <domain ...> <set><i>`
      $a_1$ `</i>...</set>`
3. Iterating over a sequence.
   **Mathematics** $\bigcup_{i=1}^{3} a_i$
   **LaTeX** `\bigcup_{i=1}^3 a_i`
   **OpenMath** `big_union` on `make_list`
   **MathML** `<apply> <union/> <bvar>i</bvar> <lowlimit>...`

It could be argued that the OpenMath is trying to mimic the LaTeX too closely, and that MathML has the right idea, that there is only one concept of 'union'. This seems to the author as being analogous to the "wishful mnemonics" issue of [13]. $\cup$ and $\bigcup$ do *not* mean the same thing: we should note that $\cup\{\{a\},\{b\}\} = \{\{a\},\{b\}\}$, while $\bigcup\{\{a\},\{b\}\} = \{a,b\}$.

In the analogy of [7], $\bigcup$ is a different part of the conjugation of $\cup$, and mathematics has more strikingly irregular verbs, so that $\Sigma$ plays the same rôle to $+$ as $\bigcup$ does to $\cup$, and indeed as $\prod$ does to juxtaposition, or MathML's `&InvisibleTimes;`.

---

[5] And therefore clearly associative and commutative. Not quite so obviously however, it should also be idempotent, since $\{a,a\} = \{a\}$. This may explain why $\bigcup\{a_1,a_2,a_3\}$ looks natural, but $\Sigma\{a_1,a_2,a_3\}$ does not.

## 4   Plus or Minus

This is familiar to us all from the solution to the quadratic:

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \tag{1}$$

which can be seen as shorthand for

$$\left\{ \frac{-b - \sqrt{b^2 - 4ac}}{2a}, \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right\}. \tag{2}$$

We are prepared to accept it in formulae such as [1, Equation 4.3.38]

$$\tan z_1 \pm \tan z_2 = \frac{\sin(z_1 \pm z_2)}{\cos z_1 \cos z_2}, \tag{3}$$

which we read as shorthand for two equations:

$$\tan z_1 - \tan z_2 = \frac{\sin(z_1 - z_2)}{\cos z_1 \cos z_2}$$

$$\tan z_1 + \tan z_2 = \frac{\sin(z_1 + z_2)}{\cos z_1 \cos z_2},$$

and the same is true of

$$\mathrm{Arctan}(z_1) \pm \mathrm{Arctan}(z_2) = \mathrm{Arctan}\left( \frac{z_1 \pm z_2}{1 \mp z_1 z_2} \right), \tag{4}$$

as meaning

$$\mathrm{Arctan}(z_1) + \mathrm{Arctan}(z_2) = \mathrm{Arctan}\left( \frac{z_1 + z_2}{1 - z_1 z_2} \right) \tag{5}$$

and

$$\mathrm{Arctan}(z_1) - \mathrm{Arctan}(z_2) = \mathrm{Arctan}\left( \frac{z_1 - z_2}{1 + z_1 z_2} \right). \tag{6}$$

But what of [1, Equations 4.6.26,27]

$$\mathrm{Arcsinh}\, z_1 \pm \mathrm{Arcsinh}\, z_2 = \mathrm{Arcsinh}\left( z_1 \sqrt{1 - z_2^2} \pm z_2 \sqrt{1 - z_1^2} \right) \tag{7}$$

$$\mathrm{Arccosh}\, z_1 \pm \mathrm{Arccosh}\, z_2 = \mathrm{Arccosh}\left( z_1 z_2 \pm \sqrt{(z_1^2 - 1)(z_2^2 - 1)} \right)? \tag{8}$$

As explained in [6], these have no such meaning, but are rather glosses on more complicated inclusions of the form $A = B \cup C$ or $A \subset B \cup C$ where $A$, $B$ and $C$ are multivalued expressions. In particular the $\pm$ on the left-hand side of (8) is redundant, since $\mathrm{Arccosh}(z) = -\mathrm{Arccosh}(z)$. (7) really means

$$\mathrm{Arcsinh}\, z_1 + \mathrm{Arcsinh}\, z_2 \subset \mathrm{Arcsinh}\left( z_1 \sqrt{1 - z_2^2} + z_2 \sqrt{1 - z_1^2} \right) \cup$$

$$\mathrm{Arcsinh}\left( z_1 \sqrt{1 - z_2^2} - z_2 \sqrt{1 - z_1^2} \right),$$

*and* the fact that the same equation holds for $\text{Arcsinh}\, z_1 - \text{Arcsinh}\, z_2$.

We are forced to conclude that $\pm$ has *no* definite meaning,

## 5   Pq

Chapter 16 of [1] is devoted to the elliptic functions such as sn, an area which has probably engendered more notational disputes and confusion as any other. [1, equation 16.25.1] defines

$$\text{Pq}(u) = \int_0^u \text{pq}^2(t)\mathrm{d}t \tag{9}$$

(where $\text{pq}^2(t)$ means $\text{pq}(t)^2$, and most certainly not $p \cdot q^2$ — see the next section). This is, of course, in defiance of (either of) the conventions of Arctan, but we are dealing with elliptic functions, not elementary ones. However, the joker here is that equation (9) applies whenever p and q are any of the letters s,c,n,d (note the order, which is traditional in the subject, and the implied assumption that $p \neq q$). Hence this equation is in fact shorthand for twelve equations of the form

$$\text{Sn}(u) = \int_0^u \text{sn}^2(t)\mathrm{d}t, \tag{10}$$

except that, when $q$ is $s$, equation (9) should be read as

$$\text{Pq}(u) = \int_0^u \left(\text{pq}^2(t) - \frac{1}{t^2}\right)\mathrm{d}t - \frac{1}{u}, \tag{11}$$

where the changes are to remove the removable singularity at $t = 0$.

A similar equation, but this time with explanation, can be seen as

$$\text{pq}(u) = \frac{\text{pr}(u)}{\text{qr}(u)} \qquad\qquad ([1, \text{Equation } 16.3.4])$$

(except that here there is no distinctness assumption, but pp is to be taken as the constant function 1).

To quote [1, coda to section 16.27]

> There is a bewildering variety of notations . . . so that in consulting books caution should be used.

As an example of this, or showing that not all apparent misprints are such, we can see [1, Equation 17.2.8–10]

$$E(u|m) = \int_0^x (1 - t^2)^{-1/2}(1 - mt^2)^{1/2}\mathrm{d}t = \int_0^u \text{dn}^2(w)\mathrm{d}w. \tag{12}$$

Does this tell us what $\text{Dn}(u)$ is — indeed [1, Equation 16.26.3] has $\text{Dn}(u) = E(u)$. However, the '$x$' in equation (12) is not a misprint, and in fact [1, Equation 17.2.2] $x = \text{sn}\, u$. So in Maple-speak

```
EllipticE(u,m)=int(sqrt((1-m*t^2)/(1-t^2)),t=0..JacobiSN(u,m))
              =int(JacobiDN(t,m)^2,t=0..u).
```

Quite how this is to be reconciled with [10, Equation 5.138(3)] —

$$\int \mathrm{dn}^2(u) = E(\mathrm{am}\, u, k)$$

— is not clear ($\mathrm{dn}(u)$ is really $\mathrm{dn}(u,m)$ of course, and $m = k^2$ here, and indeed throughout the theory, to the point where it appears to be improper to use any other letter).

## 6   The Meaning of Juxtaposition

Juxtaposition is a well-known trick notation in mathematics. It is normally believed to have two meanings (the first two listed below), but in fact has more. MathML-Presentation writes [4, 3.2.5.5] as follows.

> Certain operators that are "invisible" in traditional mathematical notation should be represented using specific entity references within `mo` elements, rather than simply by nothing.

**Multiplication.** A typical example would be $ab$, which could otherwise be rendered as $a \cdot b$. This is correctly encoded as `&InvisibleTimes;` in MathML. We should note that this only applies to italic letters, juxtaposed roman letters are deemed to constitute a single lexeme, as in sin or pq (see (9)).

**(Function) Application.** A typical example would be $\sin x$, which could otherwise be rendered as $\sin(x)$, though even in this case there is ambiguity, since $\sin(x + y)$ is different from $2(x + y)$, and $f(x + y)$ is harder to understand. This is correctly encoded as `&ApplyFunction;` in MathML.

**Concatenation.** A typical example would be $m_{12}$, which could otherwise be rendered as $m_{1,2}$. This is correctly encoded as `&InvisibleComma;` in MathML. Even without this, MathML is less ambiguous than ordinary notation: $m_{12}$ might equally be the twelfth item of a vector, but MathML would distinguish between

```
<msub>
  <mi> m </mi>
  <mrow>
    <mn> 1 </mn>
    <mn> 2 </mn>
  </mrow>
</msub>
```

and

```
<msub>
  <mi> m </mi>
```

```
    <mrow>
      <mn> 12 </mn>
    </mrow>
  </msub>
```

(of course, the `<mrow>` is redundant in the latter case.

**Addition.** A typical example would be $4\frac{1}{2}$, which could otherwise be rendered as $4 + \frac{1}{2}$. This is correctly encoded as `&InvisiblePlus;` in MathML-3[6].

**Summation.** A typical example would be $a^i b_i$, which could otherwise be rendered as $\sum_i a^i b_i$. This is the summation convention, also called Einstein notation after its introducer [9, p. 781]. It has no MathML counterpart, nor would it be easy to see how to add one. This is, of course, not so much juxtaposition in the strict sense as 'proximity'.

## 7   Letters and Fonts

We have already seen that font can make a difference in what a compiler-writer would think of as the *lexing* of mathematics: thus 'pq' is a single token, whereas '$pq$' is "$p$ juxtaposed with $q$", which might become $p$ `&InvisibleTimes;` $q$. Such lexing is already present in MathML(-Presentation) and (properly written) LaTeX: it only becomes an issue in areas such as the OCR of existing mathematics.

It is common to use fonts for semantic purposes, e.g. $x$ might be the length of the vector $\mathbf{x}$ etc. Such conventions tend to be explained (in natural language) at the start of papers, but present a real problem to a parser, which would essentially have to convert $x$ into $|\mathbf{x}|$ for internal purposes. Some authors also use change of case this way, as with $a = \det(A)$ etc. Fortunately for us, few authors go as far as [15]:

> Throughout this course, upper-case roman letters denote fields, and lower-case roman letters elements of the corresponding fields. Upper case fraktur letters denote the corresponding Galois groups, and lower case fraktur letters denote elements of the corresponding Galois groups.

## 8   Conclusion

We conclude that trying to make formal sense of natural typography can be *helped* by knowledge of both linguistic and discipline context, but that in many cases there is no obvious road to understanding questions such as whether $0 \in \mathbf{N}$ (the author probably knows), which branch cuts are intended (one hopes the author knows), or what is intended by $\pm$.

In the case of section 5 we clearly have a case of meta-notation. It is the author's contention that, in many cases (possibly even all), $\pm$ is really also

---

[6] Apparently added after several comments by this author.

metanotation, meaning "as appropriate with different choices of the signs". If the reader has a better theory, let it be explained!

Less this seems too pessimistic, we should state that, within a *given* corpus, it seems to be possible to do far better: authors generally do not change their minds wilfully during a paper ([11] is an unfortunate counter-example, dipping in and out of the summation convention several times in the course of one paper).

# References

1. Abramowitz, M., Stegun, I.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. US Government Printing Office (1964); (We quote from the tenth printing (1970): there are subtle, but occasionally significant, changes between printings)
2. Anonymous. Wikipedia, Français (2007), `http://fr.wikipedia.org`
3. Bourbaki, N.: Eléments de Mathématiques: Algèbre. C.C.L.S., Paris (1970)
4. World-Wide Web Consortium. Mathematical Markup Language (MathML) Version 3.0. W3C Working Draft (2008),
   `http://www.w3.org/TR/2007/WD-MathML3-20080409`
5. Corless, R.M., Davenport, J.H., Jeffrey, D.J., Watt, S.M.: According to Abramowitz and Stegun. SIGSAM Bulletin 2 34, 58–65 (2000)
6. Davenport, J.H.: MKM from book to computer: a case study. In: Proceedings Mathematical Knowledge Management 2003, pp. 17–29 (2003)
7. Davenport, J.H.: OpenMath in a Semantic Web,
   `http://www.jem-thematic.net/node/592`
8. Davenport, J.H., Libbrecht, P.: The Freedom to Extend OpenMath and its Utility. Mathematics in Computer Science (to appear, 2008)
9. Einstein, A.: Die Grundlage der allgemeinen Relativitaetstheorie (The Foundation of the General Theory of Relativity). Annalen der Physik Fourth Ser. 49, 284–339 (1916)
10. Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series and Products, Jeffrey A. (ed.) 5th edn. Academic Press, London (1994)
11. Lawley, D.N.: A General Method for Approximating to the Distribution of Likelihood Ratio Criteria. Biometrika 43, 295–303 (2003)
12. McDermott, D.: Artificial Intelligence Meets Natural Stupidity. SIGART Newsletter 57, 4–9 (1976)
13. McDermott, D.: Artificial Intelligence Meets Natural Stupidity. Mind Design, 143–160 (1981)
14. Oxford University Press. Oxford English Dictionary (2008),
   `http://dictionary.oed.com/entrance.dtl`
15. Roseblade, J.E.: Galois Theory. Camridge University Lectures, Michaelmas Term (1973)