# Free compact 2-categories

ANNE PRELLER and JOACHIM LAMBEK

Mathematical Structures in Computer Science / Volume 17 / Issue 02 / April 2007, pp 309 - 340
DOI: 10.1017/S0960129506005901, Published online: 18 January 2007

*Math. Struct. in Comp. Science* (2007), *vol.* 17, *pp.* 309–340. 
doi:10.1017/S0960129506005901 Printed in the United Kingdom

# Free compact 2-categories

ANNE PRELLER† and JOACHIM LAMBEK‡

†*LIRMM/CNRS Montpellier, France*
*Email:*preller@lirmm.fr

‡*McGill University Montréal, QC, Canada*
*Email:* lambek@math.mcgill.ca

*Received 10 October 2005; revised 15 May 2006*

Before one can attach a meaning to a sentence, one must distinguish different ways of parsing it. When analysing a language with pregroup grammars, we are thus led to replace the free pregroup by a free compact strict monoidal category. Since a strict monoidal category is a 2-category with one 0-cell, we investigate the free compact 2-category generated by a given category, and describe its 2-cells as labelled transition systems. In particular, we obtain a decision procedure for the equality of 2-cells in the free compact 2-category.

## 1. Introduction

An algebraic notion that has recently been applied in mathematical and computational linguistics is that of a *pregroup* (Lambek 1999), a partially ordered monoid in which each element $a$ has both a *left adjoint* $a^\ell$ and a *right adjoint* $a^r$, such that

$$a^\ell a \longrightarrow 1 \longrightarrow a^\ell a, \; aa^r \longrightarrow 1 \longrightarrow a^r a,$$

where the arrow denotes the partial order.

As a first approximation, one has recourse to the free pregroup generated by a partially ordered set of basic types. For example, consider the following English phrases:

$$\begin{array}{l} \textit{men and women} \\ \mathbf{p} \;\; \mathbf{p}^r \, \mathbf{p} \mathbf{p}^\ell \;\; \mathbf{p} \end{array} \longrightarrow \mathbf{p}$$

$$\begin{array}{l} \textit{women whom I liked} \\ \mathbf{p} \;\; \mathbf{p}^r \mathbf{p} \mathbf{o}^{\ell\ell} \mathbf{s}^\ell \;\; \pi_1 \pi^r \mathbf{s}_2 \mathbf{o}^\ell \\ \mathbf{p}\, \mathbf{p}^r \;\; \mathbf{p}\; \mathbf{o}^{\ell\ell} \; \mathbf{s}^\ell \pi_1 \pi^r \; \mathbf{s}_2 \mathbf{o}^\ell \end{array} \longrightarrow \mathbf{p}.$$

Here we have employed the following basic types:

$\pi_1$ first person subject

$\pi$ subject when the person does not matter

$\mathbf{s}_2$ sentence in the past tense

**s** sentence when tense does not matter

**p** plural noun phrase.

We also postulate

$$\mathbf{s}_2 \longrightarrow \mathbf{s},\ \pi_1 \longrightarrow \pi$$

to determine the partial order among basic types, so that, for example,

$$\pi_1 \pi^r \longrightarrow \pi\pi^r \longrightarrow 1,\ \mathbf{s}^\ell \mathbf{s}_2 \longrightarrow \mathbf{s}^\ell \mathbf{s} \longrightarrow 1.$$

Note that we have assigned to each English word a *type*, namely a string of *simple types* of the form $\cdots \mathbf{a}^{\ell\ell},\ \mathbf{a}^\ell,\ \mathbf{a},\ \mathbf{a}^r,\ \mathbf{a}^{rr} \cdots$ where **a** is any basic type. In the above example, *men* and *women* have been assigned basic types whereas

*liked* : $\pi^r \mathbf{s}_2 \mathbf{o}^\ell$

*and* : $\mathbf{p}^r \mathbf{p} \mathbf{p}^\ell$

*whom* : $\mathbf{p}^r \mathbf{p} \mathbf{o}^{\ell\ell} \mathbf{s}^\ell$.

Then

*men and* (*women whom I liked*)

$$\mathbf{p}\mathbf{p}^r\ \mathbf{p}\ \mathbf{p}^\ell \mathbf{p}\mathbf{p}^r \mathbf{p}\ \mathbf{o}^{\ell\ell} \mathbf{s}^\ell \pi_1 \pi^r\ \mathbf{s}_2\ \mathbf{o}^\ell \longrightarrow \mathbf{p}$$

(*men and women*) *whom I liked*

$$\mathbf{p}\mathbf{p}^r\ \mathbf{p}\ \mathbf{p}^\ell \mathbf{p}\mathbf{p}^r \mathbf{p}\ \mathbf{o}^{\ell\ell} \mathbf{s}^\ell \pi_1 \pi^r\ \mathbf{s}_2\ \mathbf{o}^\ell \longrightarrow \mathbf{p}.$$

These two derivations have evidently different meanings. This suggests that we should take the arrow to denote not just derivability, but the actual derivation. In other words, we should adopt the categorical imperative: replace partially ordered sets by categories. There are two distinct derivations

$$\mathbf{p}\ \mathbf{p}^r\ \mathbf{p}\ \mathbf{p}^\ell\ \mathbf{p}\ \mathbf{p}^r\ \mathbf{p} \longrightarrow \mathbf{p},$$

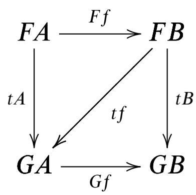
which might be thought of as morphisms in a certain category, or even, as we shall see, as 2-cells in a 2-category. Adjoints are usually defined in the 2-category of all (small) categories, but the same definition works in any 2-category. A 2-category is said to be *compact* if every 1-cell has both a left and a right adjoint.

Our interest thus shifts to compact 2-categories (originally with one 0-cell) generated by a given partially ordered set. We may as well replace this partially ordered set by a category, and we will ultimately abandon the assumption that there is only one 0-cell. Thus, we aim to study the free compact 2-category generated by a given category (or a given 2-graph).

A 2-category with one 0-cell is usually called a *strict monoidal category*. To start with, we will construct a compact one, the category of *transitions*, and show that it is equivalent to the freely generated compact strict monoidal category. The 2-cells of the category of transitions are described as what is known in computer science as labelled transition systems. Horizontal composition models parallelism; vertical composition models temporal composition of transition systems (Eilenberg 1972). Our transitions systems are given in normal form, that is, they have initial and final, but no intermediary states. Putting this another way, the 2-cells can be generated without vertical composition. The fact that every 2-cell is equal to a 2-cell in normal form is the categorical version of what logicians call 'cut-elimination'. Our proof of this fact also provides a decision procedure for the equational theory of compact 2-categories.

## 2. *2*-categories recalled

As a reminder of the concept of a 2-category, recall the notion of a natural transformation $t : F \longrightarrow G$ between functors $F : \mathbf{M} \longrightarrow \mathbf{Q}$, $G : \mathbf{M} \longrightarrow \mathbf{Q}$. Here the categories $\mathbf{M}$ and $\mathbf{Q}$ are the 0-cells, $F$ and $G$ the 1-cells and $t$ is a 2-cell. The usual definition of natural transformations requires the commutativity of the following diagram, where $f : A \longrightarrow B$ is a given arrow in the category $\mathbf{M}$:

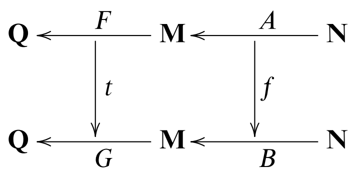



That is, the equality

$$tB \circ Ff = Gf \circ tA = tf, \text{ for } f : A \longrightarrow B, \ t : F \longrightarrow G, \tag{2.1}$$

where $\circ$ denotes the composition of 2-cells. It is reasonable to denote the diagonal by $tf$.

Now, this equality remains valid if $A$ and $B$ are themselves 1-cells, say functors $\mathbf{N} \longrightarrow \mathbf{M}$, and then $tf$ denotes *horizontal* composition $tf : FA \longrightarrow GB$ as illustrated by the diagram



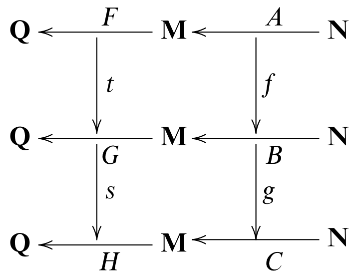

This horizontal composition is to be distinguished from the *vertical* composition

$$s \circ t : \ F \xrightarrow{t} G \xrightarrow{s} H,$$

which is the usual composition of 2-cells. The two compositions are related by the equation

$$(s \circ t)(g \circ f) = sg \circ tf, \tag{2.2}$$

which is Mac Lane's so-called *interchange law* (Mac Lane 1971).




If we identify $B$ with $1_B$ and $F$ with $1_F$, we see that (2.1) is a special case of (2.2). But (2.2) can also be deduced from (2.1) and the *distributive laws*

$$(s \circ t)C = sC \circ tC, \; F(g \circ f) = Fg \circ Ff, \tag{2.3}$$

as may be verified by diagram chasing.

As a consequence of (2.1), note that

$$1_{FA} = 1_F 1_A = 1_F A \circ F1_A = F1_A \circ 1_F A. \tag{2.4}$$

Identifying (the 2-cell) $1_F$ with (the 1-cell) $F$, (2.4) becomes

$$FA \circ FA = FA, \tag{2.5}$$

and, in the case where $A$ is an identity for horizontal composition, $F \circ F = F$. In the particular case where $f$ is the identity of the 1-cell $A$, (2.2) becomes

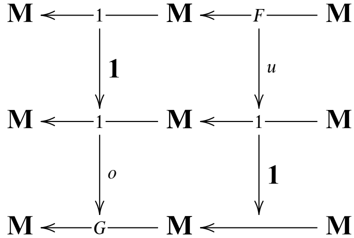
$$(s \circ t)g = (s \circ t)(g \circ 1_A) = sg \circ t1_A = sg \circ tA. \tag{2.6}$$

Finally, for $F : M \longrightarrow M,\; G : M \longrightarrow M,\; u : F \longrightarrow 1_M$ and $o : 1_M \longrightarrow G$

$$uo = ou. \tag{2.7}$$

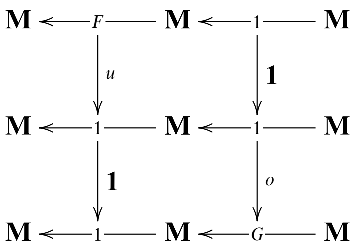
Indeed, letting $\mathbf{1}$ stand for $1_{1_{\mathbf{M}}}$ and 1 for $1_{\mathbf{M}}$, we have

$$ou = (o \circ \mathbf{1})(\mathbf{1} \circ u) = o\mathbf{1} \circ \mathbf{1}u = o \circ u$$




and, similarly,

$$uo = (\mathbf{1} \circ u)(o \circ \mathbf{1}) = \mathbf{1}o \circ u\mathbf{1} = o \circ u$$




using (2.2) and $\mathbf{1}f = f = f\mathbf{1}$.

## 3. Adjoints in 2-categories

A 1-cell $G$ is said to be a *right adjoint* of a 1-cell $F$, or $F$ a *left adjoint* of $G$, if there are 2-cells $\varepsilon : FG \longrightarrow 1$ and $\eta : 1 \longrightarrow GF$ such that

$$G\varepsilon \circ \eta G = 1_G, \qquad \varepsilon F \circ F\eta_G = 1_F$$

$$G \leftarrow GFG \leftarrow G, \quad F \leftarrow FGF \leftarrow F,$$

or, identifying $1_G$ with $G$,

$$G\varepsilon_G \circ \eta G = G, \qquad \varepsilon F \circ F\eta = F.$$

As in linguistic applications, it may be useful to call the co-unit of the adjunction $\varepsilon$ a *contraction* and the unit $\eta$ an *expansion*, and paraphrase the equations above by saying that an expansion is cancelled by a contraction immediately following it.

All the usual properties of adjoints familiar from the category of (small) categories remain valid in any 2-category. For example, adjoints are unique up to isomorphism (see, for example, (Lambek 2004)). This implies, in particular, that one can choose canonical representatives

$$G^\ell = F, \ \varepsilon_G : G^\ell G \longrightarrow 1, \ \eta_G : 1 \longrightarrow GG^\ell$$

such that

$$\begin{aligned} &G\varepsilon_G \circ \eta_G G = 1_G, \qquad \varepsilon_G G^\ell \circ G^\ell \eta_G = 1_{G^\ell} \\ &G \leftarrow GG^\ell G \leftarrow G, \quad G^\ell \leftarrow G^\ell G G^\ell \leftarrow G^\ell. \end{aligned} \tag{3.1}$$

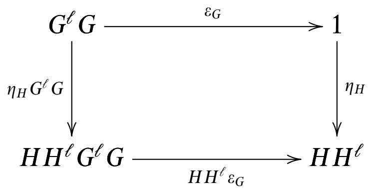
Then $G^{\ell r} \cong G \cong G^{r\ell}$, and in the category $\mathrm{T}(\mathscr{C})$ described in Section 4, these isomorphisms are replaced by the equalities

$$G^{\ell r} = G = G^{r\ell}. \tag{3.2}$$

Note that if $H$ has a left adjoint $H^\ell$ with counit $\varepsilon_H$ and unit $\eta_H$, then $GH$ has a left adjoint $H^\ell G^\ell$ with counit $\varepsilon_{GH}$ and unit $\eta_{GH}$ given by

$$\varepsilon_{GH} = \varepsilon_H \circ H^\ell \varepsilon_G H, \ \eta_{GH} = G\eta_H G^\ell \circ \eta_G. \tag{3.3}$$

Indeed, by (2.1), the diagram




commutes, so

$$\begin{aligned} GH\varepsilon_{GH} \circ \eta_{GH} GH &= GH\varepsilon_H \circ GHH^\ell \varepsilon_G H \circ G\eta_H G^\ell GH \circ \eta_G GH \\ &= GH\varepsilon_H \circ G(HH^\ell \varepsilon_G \circ G\eta_H G^\ell G)H \circ \eta_G GH \\ &= GH\varepsilon_H \circ G(\eta_H \circ \varepsilon_G)H \circ \eta_G GH \\ &= G(H\varepsilon_H \circ \eta_H H) \circ (G\varepsilon_G \circ \eta_G G)H \\ &= GH \circ GH \\ &= GH, \text{ by (2.5).} \end{aligned}$$

Similarly, $\varepsilon_{GH} H^\ell G^\ell \circ H^\ell G^\ell \eta_{GH} = H^\ell G^\ell$.

In particular, it follows that we may take

$$(GH)^\ell = H^\ell G^\ell \text{ and } (GH)^r = H^r G^r. \tag{3.4}$$

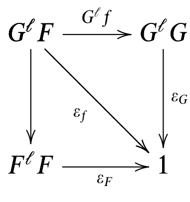
For any 2-cell $f : F \longrightarrow G$, one can define a 2-cell $f^\ell : G^\ell \longrightarrow F^\ell$ as follows:

$$f^\ell = \varepsilon_G F^\ell \circ G^\ell f F^\ell \circ G^\ell \eta_F \tag{3.5}$$

where on the right-hand side, read from right to left, the arrows are

$$F^\ell \leftarrow G^\ell G F^\ell \leftarrow G^\ell F F^\ell \leftarrow G^\ell.$$

We note that $f^\ell : G^\ell \longrightarrow F^\ell$ is the unique 2-cell that makes the following square commute:




Indeed, introducing the name *generalised contraction* for the diagonal $\varepsilon_f$, we show

$$\varepsilon_f = \varepsilon_G \circ G^\ell f = \varepsilon_F \circ f^\ell F \tag{3.6}$$

as follows:

$$\begin{aligned} \varepsilon_F \circ f^\ell F &= \varepsilon_F \circ (\varepsilon_G F^\ell \circ G^\ell f F^\ell \circ G^\ell \eta_F)F \\ &= \varepsilon_F \circ (\varepsilon_G \circ G^\ell f) F^\ell F \circ G^\ell \eta_F F \\ &= (\varepsilon_G \circ G^\ell f) \circ G^\ell F \varepsilon_F \circ G^\ell \eta_F F, \text{ by (2.1)} \\ &= \varepsilon_G \circ G^\ell f \circ G^\ell (F\varepsilon_F \circ \eta_F F) \end{aligned}$$

$$= \varepsilon_G \circ G^\ell f \circ G^\ell F$$
$$= \varepsilon_G \circ G^\ell f.$$

To show uniqueness, that is,

$$\text{If } g : G^\ell \longrightarrow F^\ell \text{ satisfies } \varepsilon_G \circ G^\ell f = \varepsilon_F \circ gF, \text{ then } g = f^\ell, \tag{3.7}$$

assume that $g$ satisfies the hypothesis. Then

$$\begin{aligned} f^\ell &= (\varepsilon_G \circ G^\ell f)F^\ell \circ G^\ell \eta_F \\ &= (\varepsilon_F \circ gF)F^\ell \circ G^\ell \eta_F \\ &= \varepsilon_F F^\ell \circ gFF^\ell \circ G^\ell \eta_F \\ &= \varepsilon_F F^\ell \circ F^\ell \eta_F \circ g, \text{ by (2.1)} \\ &= g. \end{aligned}$$

Similarly, we may define $f^r : G^r \longrightarrow F^r$ by

$$f^r = F^r \varepsilon_{G^r} \circ F^r f G^r \circ \eta_{F^r} G^r \tag{3.8}$$

and, on the way to showing uniqueness, check that it satisfies

$$f^r G \circ \eta_{G^r} = F^r f \circ \eta_{F^r}. \tag{3.9}$$

It follows that

$$f^{r\ell} = f = f^{\ell r} \tag{3.10}$$

and

$$fF^\ell \circ \eta_F = Gf^\ell \circ \eta_G = \eta_f, \tag{3.11}$$

where the *generalised expansion* $\eta_f$ is introduced as an abbreviation.

$$(g \circ f)^\ell = f^\ell \circ g^\ell, \ (g \circ f)^r = f^r \circ g^r. \tag{3.12}$$

For example, to prove $f = f^{r\ell} : F^{r\ell} \longrightarrow G^{r\ell}$, it suffices to show that $\varepsilon_{F^r} \circ F^{r\ell} f^r = \varepsilon_{G^r} \circ f G^r$, using (3.7) with $f^r : G^r \longrightarrow F^r$ instead of $f$. This can be verified as follows:

$$\begin{aligned} \varepsilon_{F^r} \circ F^{r\ell} f^r &= \varepsilon_{F^r} \circ F(F^r \varepsilon_{G^r} \circ F^r f G^r \circ \eta_{F^r} G^r) \\ &= (\varepsilon_{F^r} \circ FF^r \varepsilon_{G^r}) \circ (FF^r f \circ F\eta_{F^r})G^r, \text{ by (2.3)} \\ &= (\varepsilon_{G^r} \circ \varepsilon_{F^r} GG^r) \circ (FF^r f \circ F\eta_{F^r})G^r, \text{ by (2.1)} \\ &= \varepsilon_{G^r} \circ (\varepsilon_{F^r} G \circ FF^r f \circ F\eta_{F^r})G^r, \text{ by (2.3)} \\ &= \varepsilon_{G^r} \circ (f \circ \varepsilon_{F^r} F \circ F\eta_{F^r})G^r, \text{ by (2.1)} \\ &= \varepsilon_{G^r} \circ fG^r, \text{ by (3.1).} \end{aligned}$$

To see (3.11), we use (3.9) with $f^\ell : G^\ell \longrightarrow F^\ell$ instead of $f : F \longrightarrow G$. Finally, we derive (3.12) by a similar argument.

Equalities (3.1) generalise to

$$H\varepsilon_f \circ \eta_g F = g \circ f \text{ and } \varepsilon_g F^\ell \circ H^\ell \eta_f = (g \circ f)^\ell. \tag{3.13}$$

For example,

$$\begin{aligned} H\varepsilon_f \circ \eta_g F &= H(\varepsilon_G \circ G^\ell) \circ (gG^\ell \circ \eta_G)F \\ &= H\varepsilon_G \circ HG^\ell f \circ gG^\ell F \circ \eta_G F \\ &= H\varepsilon_G \circ gG^\ell G \circ GG^\ell f \circ \eta_G F \\ &= g \circ \varepsilon_G G \circ G\eta_G \circ f \\ &= g \circ f. \end{aligned}$$

Note that $\varepsilon_F = \varepsilon_{1_F}$, so (3.1) is a particular case of (3.13).

## 4. Transitions

A 2-category is said to be *compact*, if every 1-cell has both a left and a right adjoint. A 2-category with only one 0-cell is also called a *strict monoidal category*. For a given category $\mathscr{C}$, we will introduce a category $\mathrm{T}(\mathscr{C})$ in which the 2-cells are labelled graphs, called *transitions*, and show that it is the compact strict monoidal category freely generated by $\mathscr{C}$. As $\mathscr{C}$ is to be embedded in the free category, the objects $A, B, ..$ of $\mathscr{C}$ are identified with 1-cells, and the arrows of $\mathscr{C}$ with 2-cells such that composition in $\mathscr{C}$ becomes vertical composition in $\mathrm{T}(\mathscr{C})$. As there is only one 0-cell, horizontal composition is defined for arbitrary 1-cells and, in view of (2.1), horizontal composition is also defined for arbitrary 2-cells. Hence, let

$$\cdots, \mathbf{A}^{(-2)}, \mathbf{A}^{(-1)}, \mathbf{A}^{(0)}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)} \cdots$$

stand for

$$\cdots, \mathbf{A}^{\ell\ell}, \mathbf{A}^{\ell}, \mathbf{A}\ \mathbf{A}^{r}, \mathbf{A}^{rr} \cdots.$$

The 1-cells of $\mathrm{T}(\mathscr{C})$ are strings

$$\Gamma = \mathbf{A}_1^{(z_1)} \cdots \mathbf{A}_n^{(z_n)},\ z_i \in,\ \mathbf{A}_i \in |\mathscr{C}|,$$

where the empty string represents the unit 1. Following pregroup terminology, 1-cells of the form $\mathbf{A}^{(z)}$ are called simple types and strings of simple types are called types. Using letters $A$ and $B$ for simple types, we refer to the integer $z$ such that $A = \mathbf{A}^{(z)}$ as the *iterator* of $A$, and to $\mathbf{A}$ as the base of $A$. We define

$$\begin{aligned} (\mathbf{A}_1^{(z_1)} \ldots \mathbf{A}_n^{(z_n)})^\ell &= \mathbf{A}_n^{(z_n-1)} \ldots \mathbf{A}_1^{(z_1-1)} \\ (\mathbf{A}_1^{(z_1)} \ldots \mathbf{A}_n^{(z_n)})^r &= \mathbf{A}_n^{(z_n+1)} \ldots \mathbf{A}_1^{(z_1+1)}. \end{aligned}$$

In particular,

$$(\mathbf{A}^{(z)})^\ell = \mathbf{A}^{(z-1)},\ (\mathbf{A}^{(z)})^r = \mathbf{A}^{(z+1)}.$$

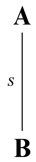
It is customary in pregroup grammars to represent contractions of simple types as under-links:

$$\varepsilon_A : A^\ell A \longrightarrow 1 \qquad\qquad \underline{A^\ell\ A}.$$

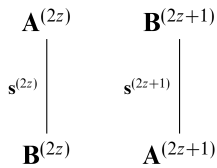
By analogy, following the practice of linear logicians, we introduce over-links for expansions of simple types:

$$\eta_A : 1 \longrightarrow AA^\ell \qquad\qquad \overline{AA^\ell}.$$

Representing an arrow $\mathbf{s} : \mathbf{A} \longrightarrow \mathbf{B}$ of $\mathscr{C}$ as a vertical link




we generalise this to vertical links




Again, $\dots, \mathbf{s}^{(-2)}, \mathbf{s}^{(-1)}, \mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots$ stands for $\dots, \mathbf{s}^{\ell\ell}, \mathbf{s}^{\ell}, \mathbf{s}\ \mathbf{s}^{r}, \mathbf{s}^{rr}, \dots$. It is convenient to declare $\mathbf{s}^{(z)} : \mathbf{A}^{(z)} \longrightarrow \mathbf{B}^{(z)}$ if either $\mathbf{s} : \mathbf{A} \longrightarrow \mathbf{B}$ and $z$ is even or $\mathbf{s} : \mathbf{B} \longrightarrow \mathbf{A}$ and $z$ is odd. We use $s : A \longrightarrow B$ for $\mathbf{s}^{(z)} : \mathbf{A}^{(z)} \longrightarrow \mathbf{B}^{(z)}$ and call arrows of this form *simple arrows*. Again, we call the integer $z$ in $s = \mathbf{s}^{(z)}$ the *iterator* of $s$ and the arrow $\mathbf{s}$ of $\mathscr{C}$ the *base of* $s$. If $s = \mathbf{s}^{(z)} : A \longrightarrow B,\ t = \mathbf{t}^{(z)} : B \longrightarrow C$, we define

$$\begin{aligned} t \circ s &= (\mathbf{t} \circ \mathbf{s})^{(z)}, \text{ if } z \text{ is even} \\ &= (\mathbf{s} \circ \mathbf{t})^{(z)}, \text{ if } z \text{ is odd.} \end{aligned}$$

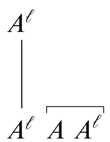
Other convenient meta-notations concerning simple arrows are

$$\begin{aligned} s^{\ell} &= (\mathbf{s}^{(z)})^{\ell} &= \mathbf{s}^{(z-1)} \\ s^{r} &= (\mathbf{s}^{(z)})^{r} &= \mathbf{s}^{(z+1)} \\ 1_{\mathbf{A}^{(z)}} &= (1_{\mathbf{A}})^{(z)}. \end{aligned}$$

It follows from these definitions that $(t \circ s)^{\ell} = s^{\ell} \circ t^{\ell}$ and $(t \circ s)^{r} = s^{r} \circ t^{r}$.

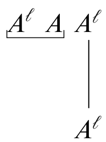
The idea is to extend this graphical representation of contractions, expansions and simple arrows to all 2-cells of the free category, using *links* labelled by simple arrows.

Horizontal composition can be represented by the juxtaposition of sets of links. For example,




represents $A^{\ell}\eta_A : A^{\ell} \longrightarrow A^{\ell}AA^{\ell}A$

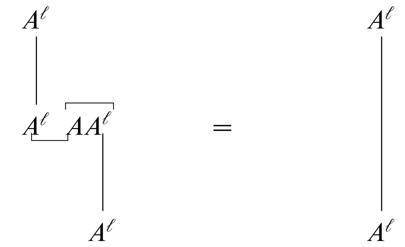
and




represents $\varepsilon_A A^{\ell} : A^{\ell}AA^{\ell}A \longrightarrow A^{\ell}$.

Vertical composition can be represented by connecting graphs vertically and identifying a composite path with the corresponding link through its endpoints. For example, for

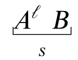
$A\varepsilon_A \circ \eta_A A = A$, we must identify



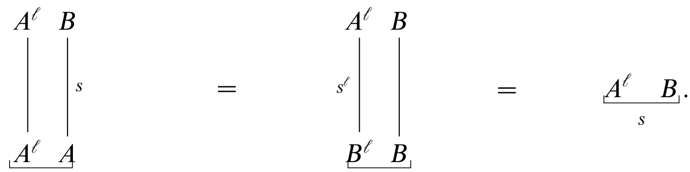
For $s : B \longrightarrow A$, we represent

$$\varepsilon_s = \varepsilon_A \circ A^\ell s = \varepsilon_B \circ s^\ell A : A^\ell B \longrightarrow 1$$

by



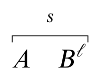
and then must define vertical composition such that



Similarly,

$$\eta_s = sB^\ell \circ \eta_B = As^\ell \circ \eta_A : 1 \longrightarrow AB^\ell$$

is represented by



In the case where the label is $\mathbf{1}_A : A \longrightarrow A$, we omit it in the graphical representation.

Prompted by the motivation above, we introduce the formal notion of a transition between strings of simple types as a special kind of graph. For the category theorist, a graph consists of two sets, the set of nodes N and the set of arrows A, and two functions, called the *domain* and *codomain*, from A to N. Graph theorists usually consider a special case of this, the so-called *directed* graph, where for each pair of nodes $(m, n)$ there is at most one arrow of domain $m$ and codomain $n$, that is, the set of arrows is identified with a binary relation on the set of nodes. Graph theorists also consider non-directed graphs, that is to say, symmetric relations, where $(m, n)$ and $(n, m)$ are identified as the *edge between* $m$ and $n$, denoted $\{m, n\}$. It is the latter version we use in the following. In fact, we will consider *labelled* non-directed graphs where a map assigns a label to each node and each edge.

**Definition 1.** Given strings of simple types $\Gamma = C_1 \cdots C_m$, $\Delta = D_1 \cdots D_n$, a *transition* $f : \Gamma \longrightarrow \Delta$ is a labelled finite non-directed graph. The nodes of $f$ have the form $(0, i)$ or $(1, k)$ where $C_i$ is the label of $(0, i)$ and $D_k$ is the label of $(1, k)$, $1 \leqslant i \leqslant m$, $1 \leqslant k \leqslant n$.

We will refer to $(0, i)$ as the 'position $i$ in the domain' of the transition and to $(1, k)$ as the 'position $k$ in the codomain'. If $i$ and $k$ are positions that are either both in the domain or both in the codomain, $i < k$ refers to the order of natural numbers. The edges, called *links* here, are divided into vertical and horizontal links, the latter being divided into over-links and under-links. The words 'vertical', 'horizontal', and so on, anticipate the graphical representation. The following must hold:

(1) A *vertical* link consists of a position $i$ in the domain and a position $k$ in the codomain. Its label is a simple arrow $s : C_i \longrightarrow D_k$.
(2) An *under*-link consists of two positions $i$ and $k$ in the domain. If $i < k$, its label is a simple arrow $s : C_k \longrightarrow C_i^r$.
(3) An *over*-link consists of two positions $i$ and $k$ in the codomain. If $i < k$, its label is a simple arrow $s : D_k^r \longrightarrow D_i$.

Moreover:

(4) Each node is an endpoint of exactly one link and every link has two distinct endpoints,
(5) If $\{(0, i), (1, k)\}$ and $\{(0, j), (1, l)\}$ are vertical links and $i < j$ in the domain, then $k < l$ in the codomain.
(6) If $\{(0, i), (0, k)\}$ is an under-link and $j$ is a position in the domain such that $i < j < k$, then $j$ belongs to an under-link $\{(0, j), (0, l)\}$ such that $i < l < k$. The same holds if we replace 'under-link' by 'over-link' and 'domain' by 'codomain'.

We will represent the transition $f : \Gamma \longrightarrow \Delta$ geometrically by a planar graph, with the domain $\Gamma = C_1 \ldots C_m$ on the top and the codomain $\Delta = D_1 \ldots D_n$ at the bottom, letting the simple types stand for their occurrences, and drawing the three kinds of links as their names indicate:

$$\begin{array}{c} C_1 \ldots C_i \ldots C_m \\ \Big|\, s \\ D_1 \ldots D_k \ldots D_n \end{array} \qquad C_1 \ldots \underbrace{C_i \ldots C_k}_{s} \ldots C_m \qquad D_1 \ldots \overbrace{D_i \ldots D_k}^{s} \ldots D_m.$$

Conditions (5) and (6) then ensure that links do not cross. If the label of a link is an identity $1_A$, we may replace it by $A$ or omit it altogether in the graphical representation.

Examples of transitions are:

— The empty graph, denoted $\mathbf{1} : 1 \longrightarrow 1$, with empty domain and empty codomain.
— For $s : A \longrightarrow B$

$$\underbrace{B^\ell \; A}_{s}$$

with domain $A^\ell \; B$ and empty codomain, which will ultimately be denoted by

$$\varepsilon_s : B^\ell A \longrightarrow 1.$$

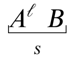
— For $t : C \longrightarrow D$

$$\overbrace{D \; C^\ell}^{t}$$

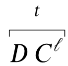
with empty domain and codomain $D\ C^{\ell}$, which will ultimately be denoted by

$$\eta_t : 1 \longrightarrow DC^{\ell}.$$

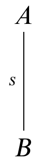
This denotation anticipates the fact that




will represent a generalised contraction and

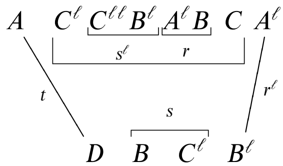



will represent a generalised expansion in the compact 2-category of transitions, as shown below. Similarly, a single vertical link




is a transition with domain $A$ and codomain $B$.

To simplify the notation, we use $s : A \longrightarrow B$ to indicate both the simple arrow and the *simple* transition with domain $A$ and codomain $B$, consisting of a unique vertical link labelled $s$. A somewhat more involved example is



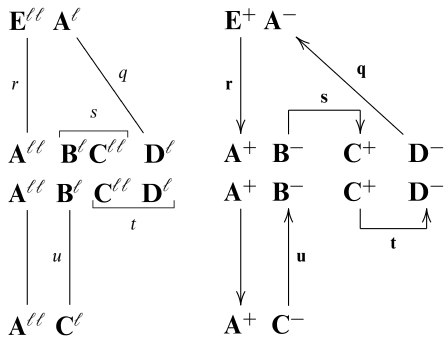

where $s : C \longrightarrow B,\ r : B \longrightarrow A,\ t : A \longrightarrow D$ and the missing labels are identities. Note that according to our notation, $s : C \longrightarrow B$ implies $s^{\ell} : B^{\ell} \longrightarrow C^{\ell} = C^{\ell\ell r}$, so $s^{\ell}$ is a correct label for a link under positions 5 and 6 in the domain above.

There is an alternative description of the labels in a transition. By definition, the label $s = \mathbf{s}^{(z)}$ of a vertical link between position $i$ in the domain and position $k$ in the codomain is a simple arrow $s : C_i \longrightarrow D_k$ and, therefore, the iterator of both $C_i$ and $D_k$ is $z$. This says that the basic arrow $\mathbf{s}$ points downward, that is, from the domain to the codomain, if $z$ is even, and upward if $z$ is odd. The label $s = \mathbf{s}^{(z)}$ of an under-link 'points' from right to left, that is, $s : C_k \longrightarrow C_i^r$ if $i < k$ in the domain. Hence, $z$ is also the iterator of $C_k$ and $z-1$ the iterator of $C_i$, that is, $C_i = \mathbf{B}^{(z-1)}$ and $C_k = \mathbf{A}^{(z)}$ and either $\mathbf{s} : \mathbf{A} \longrightarrow \mathbf{B}$ if $z$ is even, or $\mathbf{s} : \mathbf{B} \longrightarrow \mathbf{A}$ if $z$ is odd. This means that in under-links, the base arrow $\mathbf{s}$ is directed from the position with the even iterator to the position with the odd iterator.

Similarly, the label $s$ of an over-link between positions $k$ and $i$ in the codomain 'points' again from right to left, that is, $s : D_k^r \longrightarrow D_i$ for $i < k$. This time the iterator $z$ of $s$ coincides with that of $D_i$, whereas the iterator of $D_k$ is $z-1$, that is, $D_i = \mathbf{B}^{(z)}$ and

$D_k = \mathbf{A}^{(z-1)}$ and either $\mathbf{s} : \mathbf{A} \longrightarrow \mathbf{B}$ if $z$ is even, or $\mathbf{s} : \mathbf{B} \longrightarrow \mathbf{A}$ if $z$ is odd. Hence, in over-links, the base arrow $\mathbf{s}$ is directed from the base with the odd iterator to the base with the even iterator.

Consider, for example, the transitions and their base graphs



where

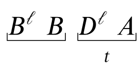
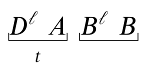
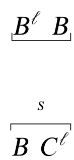
$$r : \mathbf{E}^{\ell\ell} \longrightarrow \mathbf{A}^{\ell\ell}, \quad r = \mathbf{r}^{\ell\ell}, \quad \mathbf{r} : \mathbf{E} \longrightarrow \mathbf{A}$$
$$q : \mathbf{A}^{\ell} \longrightarrow \mathbf{D}^{\ell}, \quad q = \mathbf{q}^{\ell}, \quad \mathbf{q} : \mathbf{D} \longrightarrow \mathbf{A}$$
$$s : \mathbf{C}^{\ell} \longrightarrow \mathbf{B}^{\ell}, \quad s = \mathbf{s}^{\ell}, \quad \mathbf{s} : \mathbf{B} \longrightarrow \mathbf{C}$$

and

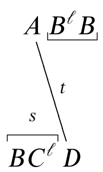
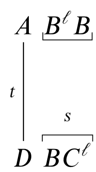
$$u : \mathbf{B}^{\ell} \longrightarrow \mathbf{C}^{\ell}, \quad u = \mathbf{u}^{\ell}, \quad \mathbf{u} : \mathbf{C} \longrightarrow \mathbf{B}$$
$$t : \mathbf{D}^{\ell} \longrightarrow \mathbf{C}^{\ell}, \quad t = \mathbf{t}^{\ell}, \quad \mathbf{t} : \mathbf{C} \longrightarrow \mathbf{D}.$$

In the right-hand graph we have replaced the links by the basic arrows, the even iterators by $+$ and the odd iterators by $-$.

We define *horizontal composition* of transitions as juxtaposition. For example, if $s : C \longrightarrow B$ and $t : A \longrightarrow D$

$\varepsilon_B\varepsilon_t =$  $\varepsilon_t\varepsilon_B =$  $\varepsilon_B\eta_s =$ 

or

$\eta_s t\varepsilon_B =$  $t\eta_s\varepsilon_B =$  .

The examples above are constructed from one-link transitions by horizontal composition, but not all transitions can be obtained in this way. Counter-examples are

$$\underbrace{B^\ell\ \underbrace{B^\ell\ B}\ \underbrace{D^\ell\ A}_{t}\ C}_{s} \qquad\qquad \overbrace{B\ \overbrace{D\ A^\ell}^{t}\ C^\ell}^{s}\ .$$

In fact, they are obtained by what we call *nesting*. We can perform nesting on transitions consisting of either under-links only or over-links only:

Let $s : A \longrightarrow B$ be a simple arrow.

- $\varepsilon_s(g) : B^\ell \Gamma A \longrightarrow 1$ is obtained from $g : \Gamma \longrightarrow 1$ by adding a new under-link from $B^\ell$ to $A$ labelled $s$
- $\eta_s(h) : 1 \longrightarrow B\Delta A^\ell$ is obtained from $h : 1 \longrightarrow \Delta$ by adding a new over-link from $B$ to $A^\ell$ labelled $s$.

With this definition, the examples above can be written as

$$\eta_s(\mathbf{1}) = \overbrace{B\ C^\ell}^{s} = \eta_s \qquad\qquad \varepsilon_t(\mathbf{1}) = \underbrace{D^\ell\ A}_{t} = \varepsilon_t$$

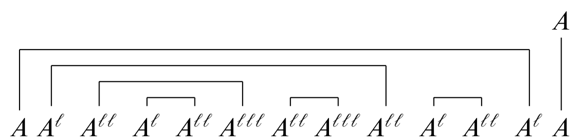
and

$$\varepsilon_s(\varepsilon_B\varepsilon_t) = \underbrace{B^\ell\ \underbrace{B^\ell\ B}\ \underbrace{D^\ell\ A}_{t}\ A}_{s} \qquad\qquad \eta_s(\eta_t(\mathbf{1})) = \overbrace{B\ \overbrace{D\ A^\ell}^{t}\ A^\ell}^{s},$$

for $s : C \longrightarrow B, \quad t : A \longrightarrow D$.

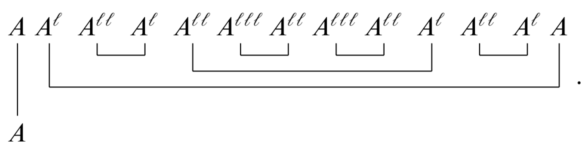
There is an obvious candidate for *vertical composition*, as we have seen in the examples at the beginning of the section, namely vertical connection of transitions where every maximal path† is replaced by the link through its endpoints. These paths can get quite involved as shown by the following example. Connect

$f \quad =$



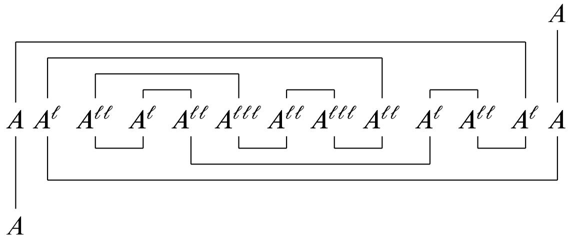

to

$g \quad =$

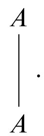



† That is, a path that has no proper extension.

The connected graph

$g;f \quad =$ 

has a unique maximal path with both endpoints labelled $A$, one in the domain and the other in the codomain

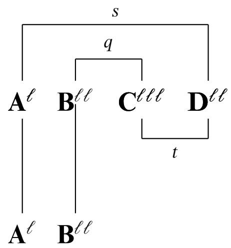
$g \circ f \quad =$  .

Note that the labels of successive links in a connected graph cannot be composed in general: starting from the right-hand upper corner, the labels of the first successive links are $1_A$ for the vertical link of $f$, $1_A$ for the longest under-link of $g$, $1_{A^\ell} = (\mathbf{1}_A)^\ell$, over-link of $f$ starting in the second position of the string, $1_{A^{\ell\ell}} = (1_A)^{\ell\ell}$, and so on. However, the base arrows of these links can be composed. Here, and below, when we say the 'iterator of a position' or 'the base of a position', we mean the iterator or the base of the simple type that is the label of the position, and similarly for links.

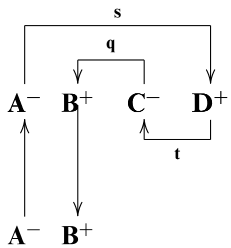
We form the *connection* $g;f$ of $f : \Gamma \longrightarrow \Delta$ *with* $g : \Delta \longrightarrow \Lambda$ *at* $\Delta$ as the union of $g$ with $f$ after renaming the nodes in the codomain of $g$ from $(1,k)$ to $(2,k)$ and those in the domain of $g$ from $(0,i)$ to $(1,i)$. Note that a maximal path in $g;f$ has its endpoints necessarily in the domain of $f$ or the codomain of $g$. We orient a maximal path as follows. A vertical path, that is, one with one endpoint in the domain of $f$ and the other in the codomain of $g$, is directed from the top (the domain of $f$) to the bottom (the codomain of $g$) if the iterator of the endpoint in the domain of $f$ is even, otherwise it is directed from the bottom to the top. If both end-points are in the domain of $f$, the path starts at the endpoint with the even iterator. If both end-points are in the codomain of $g$, it starts at the endpoint with the odd iterator. We assign a label to each path in $g;f$ as the simple arrow whose base is obtained by composing the base arrows of the successive links beginning at the starting point of the path. The iterator of the label of a vertical path is that of the starting position. If a path has both endpoints in the domain of $f$, the iterator of its label is that of its rightmost endpoint. If it has both endpoints in the codomain of $g$, the iterator of its label is that of its leftmost endpoint.

To define the *vertical composition* $g \circ f : \Gamma \longrightarrow \Lambda$ of $f : \Gamma \longrightarrow \Delta$ and $g : \Delta \longrightarrow \Lambda$, we connect $f$ with $g$ at $\Delta$ to obtain $g;f$. The links of $g \circ f$ are obtained by replacing each maximal path of $g;f$ by a single link through its endpoints. The label of the link consists of the base and the iterator of the replaced path.

To motivate the definition of the label, recall our alternative description of the labels of links. It then becomes obvious that the basic arrows along a path can be composed as indicated. For example, the connected graph




yields the base graph




where

$$\begin{array}{lll} s:\mathbf{D}^{\ell}\longrightarrow\mathbf{A}^{\ell}, & s=\mathbf{s}^{\ell}, & \mathbf{s}:\mathbf{A}\longrightarrow\mathbf{D} \\ t:\mathbf{C}^{\ell\ell}\longrightarrow\mathbf{B}^{\ell\ell}, & t=\mathbf{t}^{\ell\ell}, & \mathbf{t}:\mathbf{C}\longrightarrow\mathbf{B} \\ q:\mathbf{D}^{\ell\ell}\longrightarrow\mathbf{C}^{\ell\ell}, & q=\mathbf{q}^{\ell\ell}, & \mathbf{q}:\mathbf{D}\longrightarrow\mathbf{C}. \end{array}$$

In this case the label is $(1_{\mathbf{B}}\circ\mathbf{q}\circ\mathbf{t}\circ\mathbf{s}\circ 1_{\mathbf{A}})^{\ell}$, which is indeed a simple arrow

$$(\mathbf{q}\circ\mathbf{t}\circ\mathbf{s})^{\ell}:\mathbf{B}^{\ell}\longrightarrow\mathbf{A}^{\ell},$$

corresponding to the transition

$$\overbrace{\mathbf{A}^{\ell}\mathbf{B}^{\ell\ell}}^{(\mathbf{q}\circ\mathbf{t}\circ\mathbf{s})^{\ell}}.$$

In the next lemma we show that in general the composite of the base arrows with the chosen iterator is an appropriate label for the link replacing the path.

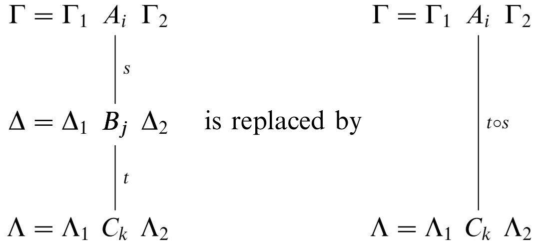
**Lemma 1 (Combing).** Let $f:\Gamma\longrightarrow\Delta$ and $g:\Delta\longrightarrow\Lambda$ be transitions, $\Gamma=A_1\cdots A_n$, $\Delta=B_1\cdots B_m$, $\Lambda=C_1\cdots C_p$. Then $g\circ f$ is a transition with domain $\Gamma$ and codomain $\Lambda$.

*Proof.* We use induction on the length $m$ of the *intermediary* string $\Delta$. If $m=0$, then $\Delta$ is empty, $f$ just has under-links, and $g$ just has over-links. Hence all paths in $g;f$ have length 1 and $g\circ f=g;f=gf$. For the induction step, assume that $\Delta$ is non-empty and that the property holds for all transitions $f':\Gamma\longrightarrow\Delta'$ and $g':\Delta'\longrightarrow\Lambda$ connected at an intermediary $\Delta'$ shorter than $\Delta$. Note that every path of length at least 2 goes through a position in $\Delta$. In the following argument, we choose a section of a path through such a position consisting of two or three consecutive links. This section will be called a *strand* and will be replaced by a single link, with the same endpoints. There are eight different strands to be considered:

**Case 1:** Suppose there is a position $j$ in $\Delta$ such that both $f$ and $g$ have a vertical link through $j$. Let $s : A_i \longrightarrow B_j$ and $t : B_j \longrightarrow C_k$ be the corresponding labels. Then $f = f_1 s f_2$ and $g = g_1 t g_2$ where $f_i : \Gamma_i \longrightarrow \Delta_i$, $g_i : \Delta_i \longrightarrow \Lambda_i$ for $i = 1, 2$. By the induction hypothesis, $g_i \circ f_i : \Gamma_i \longrightarrow \Lambda_i$ is a transition, for $i = 1, 2$, so

$$g \circ f = (g_1 \circ f_1)(t \circ s)(g_2 \circ f_2).$$

(Strand 1)




**Case 2:** (Strand 2.1) to (Strand 2.6)

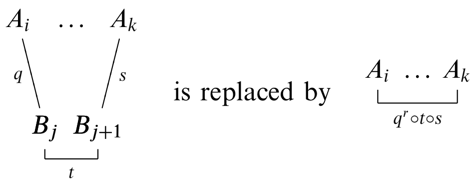
If $\Delta$ does not have such a position, assume first that $g$ has at least one under-link. Then there is a position $j$ in $\Delta$ such that $j$ and $j + 1$ form an under-link of $g$. Let $\Delta'$ be obtained from $\Delta$ by omitting $B_j B_{j+1}$ and $g'$ from $g$ by omitting the under-link through $j$ and $j + 1$. Clearly, $g'$ is a transition from $\Delta'$ to $\Lambda$. Next, consider the links determined by the positions $j$ and $j + 1$ in the codomain of $f$, say$\{(\gamma, i), (1, j)\}$ and $\{(1, j + 1), (\delta, k)\}$, where $\gamma, \delta \in \{0, 1\}$. Note that two consecutive positions $j$, $j + 1$ in $\Delta$ cannot simultaneously form an over-link of $f$ and an under-link of $g$. Indeed, the former would imply that the iterator of $B_j$ is greater than the iterator of $B_{j+1}$, whereas the latter would imply the contrary. Hence, $i$ and $k$ are both different from $j$ and $j+1$. We obtain $f'$ from $f$ by omitting the two links $\{(\gamma, i), (1, j)\}$ and $\{(1, j + 1), (\delta, k)\}$ and adding the new link $\{(\gamma, i), (\delta, k)\}$. For each strand, we verify that the labels (or their adjoints) of the three consecutive links can be composed, thus providing the label for $\{(\gamma, i), (\delta, k)\}$. Then the maximal paths of $g; f$ are identified with the maximal paths of $g'; f'$. Hence, by definition, $g \circ f = g' \circ f'$. The property then follows by the induction hypothesis.

The under-link from $B_j$ to $B_{j+1}$ being fixed in the next 6 cases, let $t : B_{j+1} \longrightarrow B_j^r$ be its label.

**Case 2.1:** Both positions $i$ and $k$ are in the domain of $f$.

As links do not cross, we have $i < k$. Let $q : A_i \longrightarrow B_j$ and $s : A_k \longrightarrow B_{j+1}$ be the labels of the corresponding vertical links. According to the notation introduced earlier, $q^r : B_j^r \longrightarrow A_i^r$, so $q^r \circ t \circ s$ is defined and is a simple arrow $q^r \circ t \circ s : A_k \longrightarrow A_i^r$.

(Strand 2.1)

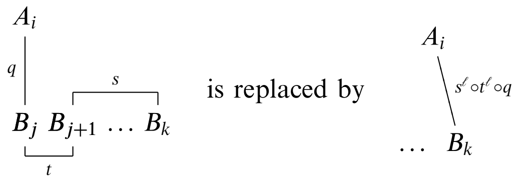



Note that the positions between $i$ and $k$ in the domain must be linked by under-links of $f$, thus defining a subtransition $f_3$ with codomain 1 of $f$. Therefore, $f = f_1 t^\ell f_3 s f_2$. Replacing the two vertical links $\{(0,i),(1,j)\}$ and $\{(0,k),(1,j+1)\}$ by a single under-link $\{(0,i),(0,k)\}$ and leaving the other links of $f$ unchanged, we obtain a transition $f'$ from $\Gamma$ to $\Delta'$.

**Case 2.2:** Position $i$ is in the domain and position $k$ in the codomain of $f$.

As links do not cross, $j+1 < k$. The label of the vertical link is a simple arrow $q : A_i \longrightarrow B_j$ and the label of the over-link is a simple arrow $s : B_k^r \longrightarrow B_{j+1}$. Then $s^\ell : {B_{j+1}}^\ell \longrightarrow B_k$ and $t^\ell : B_j \longrightarrow B_{j+1}^\ell$, so $s^\ell \circ t^\ell \circ q : A_i \longrightarrow B_k$. Hence:

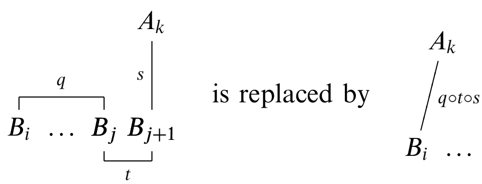
(Strand 2.2)




**Case 2.3:** Position $i$ is in the codomain and position $k$ in the domain of $f$.

As links do not cross, $i < j$. Then $q : B_j^r \longrightarrow B_i$, $s : A_k \longrightarrow B_{j+1}$ and $q \circ t \circ s : A_k \longrightarrow B_i$.

(Strand 2.3)

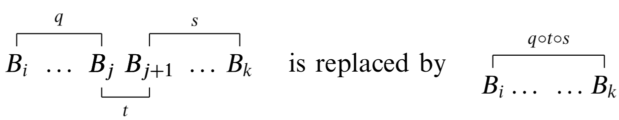



**Case 2.4:** Both positions $i$ and $k$ are in the codomain of $f$.

**Case 2.4.1:** $i < j$ and $j+1 < k$.

Let $q$ be the label of the over-link between $i$ and $j$, and $s$ be the label of the over-link between $j+1$ and $k$. Then $q : B_j^r \longrightarrow B_i$, $s : B_k^r \longrightarrow B_{j+1}$, so $q \circ t \circ s : {B_k}^r \longrightarrow B_i$.

(Strand 2.4.1)

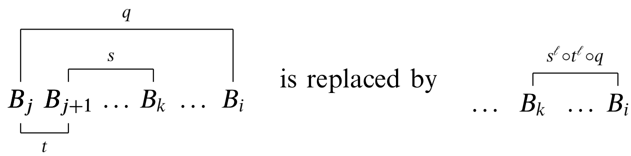


Note that the positions between $i$ and $j$ are linked by over-links in $f$, and similarly for the positions between $j+1$ and $k$. Hence $f'$ is again a transition from $\Gamma$ to $\Delta'$.

**Case 2.4.2:** $j < i$ and $j+1 < k$.

As links do not cross, it follows that $k < i$. The label of the over-link between $i$ and $j$ is a simple arrow $q : B_i^r \longrightarrow B_j$. The label of the over-link between $j+1$ and $k$ is a simple arrow $s : B_k^r \longrightarrow B_{j+1}$, so $s^\ell : B_{j+1}^\ell \longrightarrow B_k$. Hence $s^\ell \circ t^\ell \circ q : B_i^r \longrightarrow B_k$.

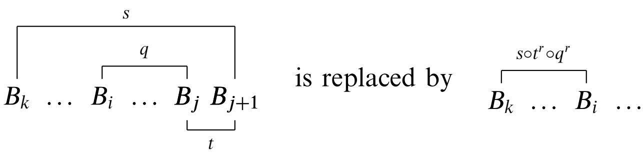
(Strand 2.4.2)




**Case 2.4.3:** $i < j$ and $k < j+1$.

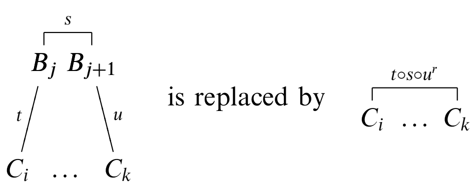
As labels, we have $q : B_j^r \longrightarrow B_i$ and $s : B_{j+1}^r \longrightarrow B_k$. Hence $s \circ t^r \circ q^r$ : $B_i^{r} \longrightarrow B_k$

(Strand 2.4.3)




**Case 3**: There remains the case where $g$ has no under-links. As we are in the case where no position in $\Delta$ belongs to both a vertical link in $g$ and a vertical link in $f$, the latter must have over-links. Hence there is a position $j$ in the codomain of $f$ linked to $j+1$ in $f$. Let $i$ and $k$ be the positions in the codomain of $g$ such that $i$ is linked to $j$ and $j+1$ to $k$ in $g$. As links do not cross, $i < k$. Then the labels of these links satisfy $s : B_{j+1}^r \longrightarrow B_j$, $t : B_j \longrightarrow C_i$, $u : B_{j+1} \longrightarrow C_k$. Therefore, $u^r : C_k^r \longrightarrow B_{j+1}^r$ and $t \circ s \circ u^r : C_k^r \longrightarrow C_i$.

(Strand 3)




This completes the proof. □

Note that the vertical composition of two transitions can be computed in time proportional to the number of links in the transitions. Indeed, it suffices to follow a maximal path exactly once, computing the label on the way as indicated in the definition.

**Proposition 1.** $T(\mathscr{C})$ is a compact strict monoidal category.

*Proof.* Vertical composition is clearly associative, the identity $1_{A_1\ldots A_n} : A_1\ldots A_n \longrightarrow A_1\ldots A_n$ consists of the obvious vertical links through corresponding simple types. The label of the link connecting position $i$ in the domain to position $i$ in the codomain is the identity of the simple type $A_i$. Recall that $\Gamma$ is identified with $1_\Gamma$. Then the equality (2.1)

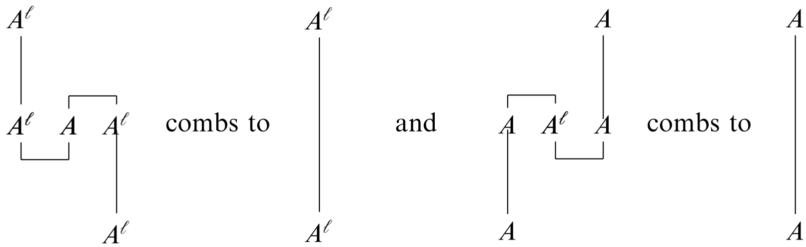
$$g\Lambda \circ \Delta f = \Theta f \circ g\Gamma = gf, \text{ for } f : \Gamma \longrightarrow \Lambda,\ g : \Delta \longrightarrow \Theta$$

is straightforward.

Compactness follows if

$$A\varepsilon_A \circ \eta_A A = A \text{ and } \varepsilon_A A^\ell \circ A^\ell \eta_A = A^\ell$$

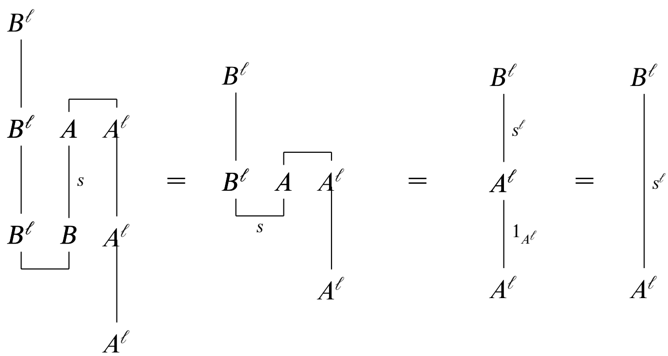
holds. By (3.3), it is enough to verify this for all simple types $A$, namely that




□

The Combing Lemma is the categorical version of cut-elimination in compact bilinear logic, which was established in Buszkowski (2002). Indeed, the categorical equality defines an equivalence relation on proofs such that transitions are cut-free representatives of equivalence classes. Besides providing a graphical representation of cut-free proofs, the categorical result tells us more: not only can we derive from $f : \Gamma \longrightarrow \Delta$ and $g : \Delta \longrightarrow \Lambda$ the existence of a cut-free $h : \Gamma \longrightarrow \Lambda$, but also show that this new $h : \Gamma \longrightarrow \Lambda$ is equivalent to $g;f$.

**Justification for the notation.** We have introduced $s^\ell = (\mathbf{s}^{(z)})^\ell = \mathbf{s}^{(z-1)}$, $s^r = (\mathbf{s}^{(z)})^r = \mathbf{s}^{(z+1)}$ for simple arrows as a convenient notation in the meta-language. Now we can show that they indeed denote the left and right adjoint, respectively, in the compact 2-category of transitions. For example, we show that $s^\ell = \varepsilon_B A^\ell \circ B^\ell s A^\ell \circ B^\ell \eta_A$:



where, from left to right, we have made the replacements (Strand 2.1), (Strand 2.2) and (Strand 1).

Similarly, 'nesting' can now be described in the language of compact 2-categories. One verifies easily that for transitions $g : \Gamma \longrightarrow 1$ and $h : 1 \longrightarrow \Delta$ and simple $s : A \longrightarrow B$,

$$\varepsilon_s(g) = \varepsilon_s \circ B^\ell g A : B^\ell \Gamma A \longrightarrow 1 \text{ and } \eta_s(h) = BhA^\ell \circ \eta_s : 1 \longrightarrow B\Delta A^\ell.$$

For example,


**Theorem 1.** $T(\mathscr{C})$ is the free compact strict monoidal category generated by $\mathscr{C}$.

*Sketch of proof.* (For a complete proof see the Appendix.) A functor $\Phi : \mathscr{C} \longrightarrow U(\mathscr{M})$ into the underlying category of another compact strict monoidal category $\mathscr{C}$, can be extended to a strict monoidal functor $\overline{\Phi} : T(\mathscr{C}) \longrightarrow \mathscr{M}$ as follows.

First we define $\overline{\Phi}$ in the obvious way on simple types and simple arrows and, writing $\overline{A}$ for $\overline{\Phi}(A)$ and $\overline{s}$ for $\overline{\Phi}(s)$, we define $\overline{\Phi}$ for generalised contractions and expansions as

$$\overline{\Phi}(\varepsilon_s) = \overline{\varepsilon_s} = \varepsilon_{\overline{s}}$$
$$\overline{\Phi}(\eta_s) = \overline{\eta_s} = \eta_{\overline{s}}.$$

Then we extend $\overline{\Phi}$ inductively to all transitions by making it commute with horizontal composition and nesting:

$$\begin{aligned} \overline{fg} &= \bar{f}\ \bar{g}\ , \\ \overline{\varepsilon_s(f)} &= \varepsilon_{\overline{s}} \circ \overline{B}^\ell \overline{f}\ \overline{A}\ , \ s : A \longrightarrow B \text{ simple,} \\ \overline{\eta_s(g)} &= \overline{B}\,\overline{g}\,\overline{A}^\ell \circ \eta_{\overline{s}}\ , \ s : A \longrightarrow B \text{ simple.} \end{aligned}$$

By construction, $\bar{\Phi}$ preserves horizontal composition, $\varepsilon$ and $\eta$. As uniqueness is obvious, it only remains to show that $\bar{\Phi}$ preserves vertical composition. To do this, we follow the Combing Lemma. For the induction step, we prove **Case 1** thus

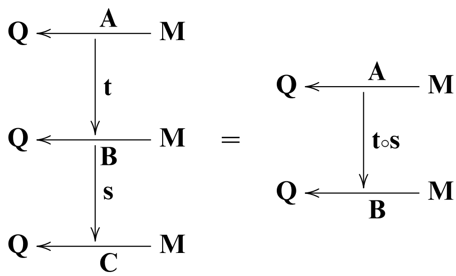
$$\begin{aligned} \overline{g \circ f} &= \overline{(g_1 \circ f_1)(t \circ s)(g_2 \circ f_2)} \\ &= \overline{(g_1 \circ f_1)}\ \overline{(t \circ s)}\ \overline{(g_2 \circ f_2)} \\ &= (\overline{g_1} \circ \overline{f_1})(\overline{t} \circ \overline{s})(\overline{g_2} \circ \overline{f_2}) \\ &= \overline{g_1}\overline{t}\overline{g_2} \circ \overline{f_1}\overline{s}\overline{f_2} \\ &= \overline{g} \circ \overline{f}. \end{aligned}$$

In the other cases we use the intermediary transitions $g'$ and $f'$ for which $g \circ f = g' \circ f'$, and therefore $\overline{g \circ f} = \overline{g' \circ f'}$ also. As, by the induction hypothesis, $\overline{g' \circ f'} = \overline{g'} \circ \overline{f'}$, it just remains to show that $\overline{g} \circ \overline{f} = \overline{g'} \circ \overline{f'}$. This requires some care as we must express the seven definitions of $g'$ and $f'$ of the Combing Lemma in the language of $T(\mathscr{C})$. Instead of carrying out the details of this program for all seven cases, a different proof will be presented in the Appendix, relating transitions to derivations in the free pregroup. □

This theorem provides a decision procedure for the equational theory of strict compact monoidal categories given by the axioms of strict monoidal categories together with (3.1) to (3.4). The procedure then also applies to any definitially equivalent theory such as that of compact non-symmetric star-autonomous categories where the unit of the tensor product is a dualising object (Barr 1995). Indeed, to decide whether $f = g$ can be derived, interpret both terms in the category of transitions.

## 5. The free strict compact 2-category generated by a given 2-graph

We can modify the above construction to the compact 2-category freely generated from a given 2-graph. To simplify matters, we will assume that the 2-cells of the 2-graph form a category:




Then the construction is the same as above. However, if $\mathbf{A} : \mathbf{M} \longrightarrow \mathbf{N}$ is a 1-cell and $z \in \mathbb{Z}$, we have to require that the simple type $\mathbf{A}^{(z)}$ is a 1-cell such that

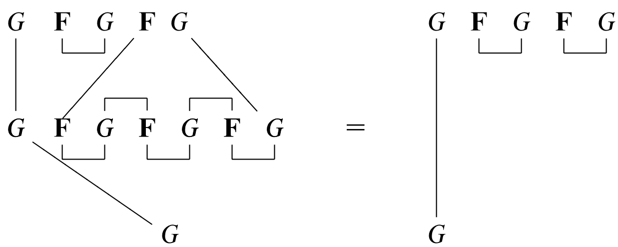
$$\mathbf{A}^{(z)} : \mathbf{M} \longrightarrow \mathbf{N}, \text{ if } z \text{ is even}$$
$$\mathbf{A}^{(z)} : \mathbf{N} \longrightarrow \mathbf{M}, \text{ if } z \text{ is odd.}$$

Types are now paths, that is, $\mathbf{A}_1^{(z_1)} \dots A_n^{(z_n)}$ must satisfy

$$\mathbf{A}_i^{(z_i)} : \mathbf{N}_i \longrightarrow \mathbf{N}_{i+1},\ 1 \leqslant i \leqslant n-1.$$

Then the 1-cells of the free compact 2-category are the types and the 2-cells are the transitions between types.

As a particular case, let $\mathscr{C}$ consist of two 0-cells, $\mathbf{M}$ and $\mathbf{N}$, a 1-cell $\mathbf{F} : \mathbf{M} \longrightarrow \mathbf{N}$ and the identity of $\mathbf{F}$ as the unique 2-cells. Let $G = \mathbf{F}^r$ and only consider transitions with domain and codomain of the form $G\mathbf{F}G \dots \mathbf{F}G$ where $\mathbf{F}G$ is repeated $n$ times, $n \geqslant 0$. Then the only possible under-links are between neighbouring $\mathbf{F}G$ in the domain and the only possible over-links between neighbouring $G\mathbf{F}$ in the codomain. Hence the first position in the domain always belongs to a vertical link. When connecting two such transitions, say



we get that Strands 2.4.2 and 2.4.3 do not occur. More generally, there is no nesting. These graphs are considered in Došen (2002) under the name of friezes. The connection between a free adjoint functor pair and cut-elimination is investigated in Došen (1999). In compact 2-categories the infinite number of adjoints requires more involved graphs for the computation of composition, like the spiral in Example 1. The 1-cells involving **F** and $\mathbf{F}^r$ only are so-called 'linear' types, see Degeilh and Preller (2005), where it was shown that there is at most one transition between two given types. In particular, linear types do not capture differences in meaning for which the presence of both right and left adjoint is required. Linguistic applications call for right and left iterated adjoints, for example, to describe the Chomskyan trace, see Lambek (1999).

One may wish to generalise the present results to bicategories, using the notions of adjunctions on bicategories (see, for example, Lambek (2004)), but we will refrain from doing so here. The special case of compact symmetric monoidal categories has been treated in the classical paper Kelly and Laplaza (1980). They did not actually construct the free such category, instead they established the important result that equations between morphisms in the language of such categories follow from the axioms if and only if they hold, up to isomorphism, for the graphs. In the situation we have discussed here, the graphs have to be equal.

## 6. Conclusion

We have described the 2-cells of $T(\mathscr{C})$, the free compact monoidal 2-category generated by $\mathscr{C}$ as labelled transition systems. These transition systems draw their labels from $\mathscr{C}$ and are closed under parallel and sequential composition. In the case where $\mathscr{C}$ is itself freely generated by a labelled graph, the edges of this graph stand for non-logical axioms or 'information'. Both left and right adjoint provide a mechanism for storing this information. It follows from the above that equality in $T\mathscr{C})$ is decidable if the equality of arrows in $\mathscr{C}$ is. This is, in particular, the case if $\mathscr{C}$ is freely generated by a labelled graph. The reductions constructed when analysing syntax with a pregroup grammar are particular transitions. As different reductions give rise to different semantical interpretations, transitions are an indispensable step from pregroup grammars to discourse representation.

## Appendix (by Anne Preller).

To prove that $T(\mathscr{C})$ is the free compact strict monoidal category, we define the extension $\overline{\Phi} : T(\mathscr{C}) \longrightarrow \mathscr{M}$ of the functor $\Phi$ from $\mathscr{C}$ to a compact strict monoidal category $\mathscr{M}$, as indicated in the outline of the proof in Section 4. First we check that $\overline{\Phi}$ is well defined. The other property left to be shown is that $\overline{\Phi}$ commutes with vertical composition. The proof outlined in Section 4 is based on the idea that the Combing Lemma can be expressed in purely categorical terms. Though the equalities corresponding to the eight cases of the Combing Lemma can be shown to hold in $\mathscr{M}$, the proof below follows a different line: it relates transitions directly to the derivations in free pregroups defined in Lambek (1999).

We noted in Section 4 that an arbitrary transition can be obtained from single links by the graphical operations of juxtaposition and nesting. To express these operations in categorical language, we distinguish the *horizontal normal forms* among the expressions of the language of compact strict monoidal categories with constants in $\mathscr{C}$.

**Definition 2 (Horizontal normal form).** Every simple arrow $s : A \longrightarrow B$, every generalised contraction $\varepsilon_s : B^\ell A \longrightarrow 1$ and every generalised expansion $\eta_s : 1 \longrightarrow BA^\ell$ is a horizontal normal form.
An arbitrary horizontal normal form is obtained from them by the following rules:

$$\text{(Horizontal composition)} \quad \frac{f : \Gamma \longrightarrow \Delta \text{ normal } g : \Theta \longrightarrow \Lambda \text{ normal}}{fg : \Gamma\Theta \longrightarrow \Delta\Lambda \text{ normal}}$$

$$\text{(Nesting Contraction)} \quad \frac{f : \Gamma \longrightarrow 1 \text{ normal } s : A \longrightarrow B \text{ simple}}{\varepsilon_s \circ B^\ell f A : B^\ell \Gamma A \longrightarrow 1 \text{ normal}}$$

$$\text{(Nesting Expansion)} \quad \frac{f : 1 \longrightarrow \Delta \text{ normal } s : A \longrightarrow B \text{ simple}}{\eta_s \circ B f A^\ell : 1 \longrightarrow B\Delta A^\ell \text{ normal}}$$

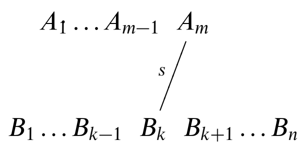
where the Horizontal Composition rule does not apply to $u : \Gamma \longrightarrow 1$ and $o : 1 \longrightarrow \Lambda$.

Note that the Horizontal Composition rule applies to $o : 1 \longrightarrow \Lambda$ and $u : \Gamma \longrightarrow 1$. The order $u : \Gamma \longrightarrow 1$ and $o : 1 \longrightarrow \Lambda$ is excluded because $uo = ou$ holds in all *2*-categories by (2.7). Thus, only $ou$ is a normal expression. This, together with the fact that **1** is not a normal expression, makes it possible to assert the uniqueness of horizontal normal forms as follows.

**Lemma 2 (Horizontal normal form).** Every non-empty transition $f : A_1 \dots A_m \longrightarrow B_1 \dots B_n$ can be expressed in horizontal normal form, which is unique up to associativity of horizontal composition.

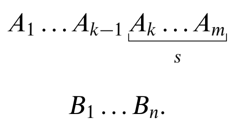
*Proof.* We use induction on the number of links in $f$. At least one of $n$ or $m$ is greater than 0. First, assume that $m > 0$. We distinguish two cases:

1 The last position $m$ of $\Gamma$ is linked to a position $k$ in the codomain $\Delta$ with label $s$:



Then the other links of $f$ can be divided into those with no endpoint to the right of $k$ and those with both endpoints to the right of $k$. The former set of links defines a transition $g : A_1 \dots A_{m-1} \longrightarrow B_1 \dots B_{k-1}$, and the latter a transition $h : 1 \longrightarrow B_{k+1} \dots B_n$ such that $f = gsh$.

2 The last position $m$ in the domain is linked to a position $k < m$ in the domain:



Let $g$ consist of the links of $f$ with endpoints in the codomain or to the left of $k$ in the domain, and let $h$ consist of the links with both endpoints in the domain strictly between $k$ and $m$. Then $g : A_1 \dots A_{k-1} \longrightarrow B_1 \dots B_n$, $h : A_{k+1} \dots A_{m-1} \longrightarrow 1$ and $f = g\varepsilon_s(h)$

Now suppose $m = 0$ and $n > 0$, and consider the link through the last position $n$ in the codomain. Let $t$ be its label. The other endpoint of this link is a position $j < n$ in the codomain:

$$B_1 \dots B_{j-1} \overset{t}{\overline{B_j \dots B_n}}.$$

Then the links that have both endpoints to the left of $j$ form a transition $g : 1 \longrightarrow B_1 \dots B_{j-1}$, and the links with both endpoints between $j$ and $n$ form a transition $h : 1 \longrightarrow B_{j+1} \dots B_{n-1}$ such that $f = g\eta_t(h)$.

From the existence of a unique normal form for a transition, it follows at once that the canonical extension $\overline{\Phi}$ is well defined. We recall the definition using $\overline{(\ )}$ instead of $\overline{\Phi}$ :

(I) $\overline{\mathbf{A}^{(0)}} = \Phi(\mathbf{A})$, $\mathbf{A}$ object of $C$
$\overline{\mathbf{s}^{(0)}} = \Phi(\mathbf{s})$, $\mathbf{s}$ arrow of $C$
(II) $\overline{\mathbf{A}^{n+1}} = \overline{\mathbf{A}^{(n)}}^r$, $\overline{\mathbf{A}^{(-n-1)}} = \overline{\mathbf{A}^{(-n)}}^\ell$, for $0 \leqslant n$
$\overline{\mathbf{s}^{(n+1)}} = \overline{\mathbf{s}^{(n)}}^r$, $\overline{\mathbf{s}^{(-n-1)}} = \overline{\mathbf{s}^{(-n)}}^\ell$, for $0 \leqslant n$
(III) $\overline{\Gamma\Delta} = \overline{\Gamma}\ \overline{\Delta}\ \overline{fg} = \overline{f}\,\overline{g}$
(IV) $\overline{\varepsilon_s} = \varepsilon_{\overline{s}}$
$\overline{\eta_s} = \eta_{\overline{s}}$
$\overline{\varepsilon_s(f)} = \varepsilon_{\overline{s}} \circ \overline{B}^\ell \overline{f}\,\overline{A} = \varepsilon_{\overline{s}}(f)$, $f : \Gamma \longrightarrow 1$, $s : A \longrightarrow B$ simple
$\overline{\eta_s(g)} = \overline{B}\,\overline{g}\,\overline{A}^\ell \circ \eta_{\overline{s}} = \eta_{\overline{s}}(g)$, $g : 1 \longrightarrow \Delta$, $s : A \longrightarrow B$ simple
(V) $\overline{1} = 1$, $\overline{1_\Gamma} = 1_{\overline{\Gamma}}$.

By definition, $\overline{\Phi}$ preserves horizontal composition and the identities. If the left and right adjoints of 1-cells are part of the signature of $\mathscr{M}$, $\overline{\Phi}$ preserves left and right adjoints only up to isomorphism in general. For example, we may just have $(GH)^\ell \cong H^\ell G^\ell$ in $\mathscr{M}$. However, as only the existence of left and right adjoints of 1-cells is assumed in the definition in Section 3, a functor of 2-categories that preserves left and right adjoint up to isomorphism may still be correctly called a functor of compact 2-categories.

Finally, we must show that $\overline{\Phi}$ commutes with vertical composition. This is easily verified if the composed transitions are simple arrows or if one of them is an identity. In the general case, the idea is to prove the property for transitions that consist essentially of just one link, the so-called *single-step* transitions, and to show that an arbitrary transition is equal to a vertical composition of single steps. □

**Definition 3 (Single step).** A *single step* is a 2-cell of one of the following forms:

| | |
|---|---|
| $\Gamma s\Delta : \Gamma A\Delta \longrightarrow \Gamma B\Delta$ | (Induced step) |
| $\Gamma\varepsilon_s\Delta : \Gamma B^\ell A\Delta \longrightarrow \Gamma\Delta$ | (Generalised contraction step) |
| $\Gamma\eta_s\Delta : \Gamma\Delta \longrightarrow \Gamma BA^\ell\Delta$ | (Generalised expansion step) |

where $s : A \longrightarrow B$ is a simple arrow.

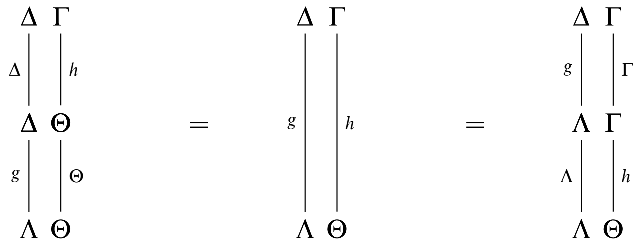
This definition uses categorical language only, so, replacing $s$ by $\bar{s}$, we may say that the canonical map preserves single steps, that is, $\overline{\Gamma s \Delta} = \overline{\Gamma}\,\bar{s}\,\overline{\Delta}$, $\overline{\Gamma \varepsilon_s \Delta} = \overline{\Gamma}\varepsilon_{\bar{s}}\overline{\Delta}$ and $\overline{\Gamma \eta_s \Delta} = \overline{\Gamma}\eta_{\bar{s}}\overline{\Delta}$. Single steps generate all transitions, as follows from Lemma 2 and the following lemma.

**Lemma 3 (Vertical decomposition of horizontal normal forms).** Every horizontal normal form $f : A_1 \ldots A_n \longrightarrow B_1 \ldots B_m$ can be expressed as a vertical composition of single steps $f = f_1 \circ \ldots \circ f_n$ such that $\bar{f} = \overline{f_1} \circ \ldots \circ \overline{f_n}$.

The proof of Lemma 3 is straightforward by induction on the derivation of the horizontal normal form of $f$. The distributivity laws (2.3) intervene if one of the nesting rules was applied. If the horizontal composition rule was applied, the argument is as follows.

For $h : \Gamma \longrightarrow \Theta$ and $g : \Delta \longrightarrow \Lambda$, the equalities

$$g\Theta \circ \Delta h = gh = \Lambda h \circ g\Gamma$$




hold in an arbitrary *2*-category by (2.1), therefore

$$\bar{g}\,\overline{\Theta} \circ \overline{\Delta}\,\bar{h} = \bar{g}\bar{h} = \overline{\Lambda}\;\bar{h} \circ \bar{g}\,\overline{\Gamma}.$$

Hence,

$$\overline{g\Theta \circ \Delta h} = \overline{gh} = \bar{g}\,\bar{h} = \bar{g}\,\overline{\Theta} \circ \overline{\Delta}\,\bar{h} = \overline{g\Theta} \circ \overline{\Delta h}$$

and, similarly,

$$\overline{\Lambda h \circ g\Gamma} = \overline{\Lambda h} \circ \overline{g\Gamma}.$$

In particular, if $h$ and $g$ are single steps, $\Lambda h$, $g\Gamma$, $g\Theta$ and $\Delta h$ are again single steps. We call $\Lambda h$ and $g\Gamma$, respectively, $g\Theta$ and $\Delta h$ *disjoint*, because the essential links cannot interact. This operation, which switches two disjoint single steps, has given the Switching Lemma of Lambek (1999) its name.

In general, however, Lemma 3 is not sufficient to show that $\overline{g \circ f} = \bar{g} \circ \bar{f}$, because $g \circ f$ is in general not in horizontal normal form. All we can conclude form this is that $g \circ f = g_1 \ldots \circ g_n \circ \circ f_1 \ldots f_m$ and that $\bar{g} \circ \bar{f} = \overline{g_1} \ldots \circ \overline{g_n} \circ \overline{f_1} \ldots \overline{f_m}$. Our next task is to associate to a vertical composition of single steps $f_1 \circ \ldots \circ f_n$ a normal form $f$ such that

$$f_1 \circ \ldots \circ f_n \;=\; f$$
$$\overline{f_1} \circ \ldots \circ \overline{f_n} \;=\; \bar{f}$$

and, therefore,

$$\overline{f_1 \circ \ldots \circ f_n} = \overline{f_1} \circ \ldots \circ \overline{f_n}.$$

The other operations introduced in the Switching Lemma imply this equality for $n = 2$ by replacing two successive single steps by one single step. We recall them as Operations (1) to (4) below and prove that the replaced steps are equal to the replacing step.

**Switching Operations**

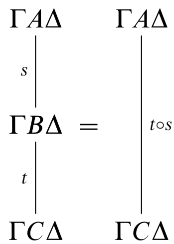
(0) Switch two disjoint steps.

$$f_i \circ f_{i+1} = f_{i+1} \circ f_i \quad \text{and} \quad \overline{f_i \circ f_{i+1}} = \overline{f_i} \circ \overline{f_{i+1}}.$$

In Operations (1) to (4) below, the two replaced step are non-disjoint.

(1) Replace two induced steps by a single induced step:

$$\Gamma t\Delta \circ \Gamma s\Delta = \Gamma(t \circ s)\Delta$$



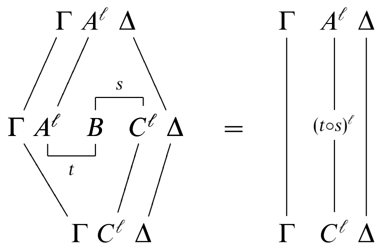
As the equality is an instance of the distributive laws in 2-categories, we also have

$$\overline{\Gamma}\,\bar{t}\,\overline{\Delta} \circ \overline{\Gamma}\,\bar{s}\,\overline{\Delta} = \overline{\Gamma}\,(\bar{t} \circ \bar{s})\overline{\Delta} = \overline{\Gamma}\ \overline{t \circ s}\ \overline{\Delta}.$$

(2) Replace a generalised expansion followed by a generalised contraction by an induced step.

(2a) The generalised contraction is on the left:

$$\Gamma\varepsilon_t C^\ell\Delta \circ \Gamma A^\ell \eta_s \Delta = \Gamma(t \circ s)^\ell \Delta$$



where $t : B \longrightarrow A$ and $s : C \longrightarrow B$. The equalities

$$\begin{aligned} \varepsilon_t C^\ell \circ A^\ell \eta_s &= (\varepsilon_A \circ A^\ell t)C^\ell \circ A^\ell(sC^\ell \circ \eta_C) \\ &= \varepsilon_A C^\ell \circ A^\ell tC^\ell \circ A^\ell sC^\ell \circ A^\ell \eta_C \\ &= \varepsilon_A C^\ell \circ A^\ell(t \circ s)C^\ell \circ A^\ell \eta_C \\ &= (t \circ s)^\ell, \text{ by (3.5)} \end{aligned}$$

and

$$\Gamma\varepsilon_t C^\ell\Delta \circ \Gamma A^\ell\eta_s\Delta = \Gamma(\varepsilon_t C^\ell \circ A^\ell \eta_s)\Delta = \Gamma(t \circ s)^\ell\Delta$$

hold in arbitrary 2-categories. Recall that the canonical map commutes with the vertical composition of simple arrows and the adjoints of simple arrows

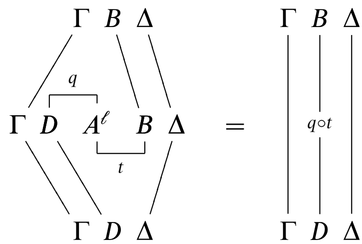
$$\overline{(t \circ s)^{\ell}} = (\bar{t} \circ \bar{s})^{\ell}.$$

Hence

$$\begin{aligned} \overline{\Gamma\varepsilon_t C^{\ell}\Delta \circ \Gamma A^{\ell}\eta_s\Delta} &= \overline{\Gamma(t \circ s)^{\ell}\Delta} \\ &= \overline{\Gamma}\ \overline{(t \circ s)^{\ell}}\ \overline{\Delta} \\ &= \overline{\Gamma}\ (\bar{t} \circ \bar{s})^{\ell}\overline{\Delta} \\ &= \overline{\Gamma}\varepsilon_{\bar{t}}\overline{\Delta} \circ \overline{\Gamma}\eta_{\bar{s}}\overline{\Delta} \\ &= \overline{\Gamma\varepsilon_t\Delta} \circ \overline{\Gamma\eta_s\Delta}. \end{aligned}$$

(2b) The generalised contraction is on the right:

$$\Gamma D\varepsilon_t\Delta \circ \Gamma\eta_q B\Delta = \Gamma(q \circ t)\Delta$$



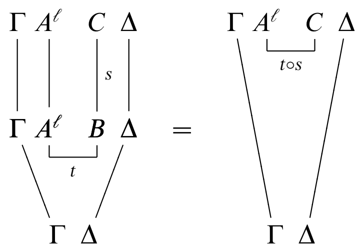

The proof is similar, using an instance of (3.13)

$$D\varepsilon_t \circ \eta_q B = q \circ t.$$

(3) Replace an induced step followed by a generalised contraction by a generalised contraction.

(3a) The essential link of the induced step is on the right

$$\Gamma\varepsilon_t\Delta \circ \Gamma A^{\ell} s\Delta = \Gamma\varepsilon_{t\circ s}\Delta$$




where $t : B \longrightarrow A$, $s : C \longrightarrow B$.
Indeed,

$$\varepsilon_t \circ A^{\ell} s = \varepsilon_A \circ A^{\ell} t \circ A^{\ell} s = \varepsilon_A \circ A^{\ell}(t \circ s) = \varepsilon_{t\circ s},$$

and

$$\Gamma\varepsilon_t\Delta \circ \Gamma A^{\ell} s\Delta = \Gamma(\varepsilon_t \circ A^{\ell} s)\Delta = \Gamma\varepsilon_{t\circ s}\Delta.$$

hold in all compact 2-categories, hence

$$\overline{\Gamma}\varepsilon_{\bar{t}}\overline{\Delta} \circ \overline{\Gamma}\,\overline{A^{\ell}}\bar{s}\overline{\Delta} = \overline{\Gamma}\varepsilon_{\bar{t}\circ\bar{s}}\overline{\Delta}.$$

(3b) The essential link of the induced step is on the left.

(4) Replace a generalised expansion and a following induced step by a generalised expansion.

(4a) The essential link of the induced step is on the right.

(4b) The essential link of the induced step is on the left.

The proofs of Cases (3b), (4a) and (4b) are left to the reader.

There are four cases that are not included in the switching operations, namely the cases where the two consecutive single steps $f_i \circ f_{i+1}$ are either both generalised contractions or both generalised expansions, or where an induced step is preceded by a generalised contraction or followed by a generalised expansion. For them also there is an intermediary transition $f$ such that

$$\begin{aligned} f_i \circ f_{i+1} &= f \\ \overline{f_i} \circ \overline{f_{i+1}} &= \overline{f}. \end{aligned}$$

However, in contrast with the cases of the switching operations (1) to (4), $f$ is not a single step but a horizontal normal form. We will prove this for a vertical composition of arbitrary length, provided the single steps are all of the same kind. The import of this property is explained by the fact that the Switching Lemma in Lambek (1999) preserves equality.

**Lemma 4 (Switching).** Every vertical composition of single steps can be rewritten as a vertical composition of single steps

$$f_1 \circ \ldots \circ f_n = (h_1 \circ \ldots \circ h_q) \circ (v_1 \circ \ldots \circ v_m) \circ (g_1 \circ \ldots \circ g_p)$$

such that the $g_i$'s are generalised contractions, the $v_i$'s induced steps and the $h_i$'s are generalised expansions. Moreover,

$$\overline{f_1} \circ \ldots \circ \overline{f_n} = (\overline{h_1} \circ \ldots \circ \overline{h_q}) \circ (\overline{v_1} \circ \ldots \circ \overline{v_m}) \circ (\overline{g_1} \circ \ldots \circ \overline{g_p}).$$

*Proof.* Omit the induced steps that are identities and use the switching operations (0) to (4). □

The horizontal normal forms corresponding to a vertical composition of single steps that are all of the same kind are described as follows.

**Definition 4.** A *normal contraction step* is a horizontal composition

$$u_0 B_1 \ldots u_{m-1} B_m u_m : \Delta_0 B_1 \ldots \Delta_{m-1} B_m \Delta_m \longrightarrow B_1 \ldots B_m$$

where $u_k : \Delta_k \longrightarrow 1$ is $\mathbf{1}$ or a horizontal normal form, with $0 \leqslant k \leqslant m$.

A *normal expansion step* is a horizontal composition

$$o_0 C_1 \ldots o_{m-1} C_m o_m : C_1 \ldots C_m \longrightarrow \Gamma_0 C_1 \ldots \Gamma_{m-1} C_m \Gamma_m$$

where $o_k : 1 \longrightarrow \Gamma_k$ is $\mathbf{1}$ or a horizontal normal form, with $0 \leqslant k \leqslant m$.

A *normal vertical step* is a horizontal composition

$$s_1 \dots s_m : B_1 \dots B_m \longrightarrow C_1 \dots C_m$$

where $s_k : B_k \longrightarrow C_k$ is a simple arrow.

We note that these normal steps generalise the single steps and are horizontal normal forms.

**Lemma 5.** Every vertical composition of generalised contractions

$$g_1 \circ \dots \circ g_p : A_1 \dots A_n \longrightarrow B_1 \dots B_m$$

can be rewritten as a normal contraction step

$$u_0 B_1 \dots u_{m-1} B_m u_m$$

such that

$$g_1 \circ \dots \circ g_p = u_0 B_1 \dots u_{m-1} B_m u_m$$

and

$$\overline{g}_1 \circ \dots \circ \overline{g}_p = \overline{u_0}\ \overline{B_1} \dots \overline{u_{m-1}}\ \overline{B_m}\ \overline{u_m}.$$

Moreover,

$$\overline{g_1 \circ \dots \circ g_p} = \overline{g}_1 \circ \dots \circ \overline{g}_p.$$

*Proof.* We use induction on the length $p$ of the vertical decomposition. Note that

$$g_1 = B_1 \dots B_j \varepsilon_t B_{j+1} \dots B_m$$

where $\varepsilon_t : A_i A_k \longrightarrow 1$ for some $1 \leqslant i < k \leqslant n$. By the induction hypothesis,

$$g_2 \circ \dots \circ g_p = f' A_i u A_k f'',$$

where $u : A_{i+1} \dots A_{k-1} \longrightarrow 1$ is the identity $\mathbf{1}$ or in normal form, and where $f' : A_1 \dots A_{i-1} \longrightarrow B_1 \dots B_j$ and $f'' : A_{k+1} \dots A_n \longrightarrow B_{j+1} \dots B_m$ are normal contraction steps. Hence,

$$\begin{aligned} g_1 \circ g_2 \circ \dots \circ g_p &= (B_1 \dots B_j \varepsilon_t B_{j+1} \dots B_m) \circ (f' A_i u A_k f'') \\ &= f'(\varepsilon_t \circ (A_i u A_k)) f'' \qquad \text{by (2.2)} \\ &= f' \varepsilon_t(u) f''. \end{aligned}$$

Recall that $f' = u_0{}' B_1 \dots u_{j-1}{}' B_j u_j{}'$ and $f'' = u_0{}'' B_{j+1} \dots B_m u_{m-j}{}''$, and define

$$\begin{aligned} u_l &= u_l{}', && \text{for } 0 \leqslant l \leqslant j-1 \\ u_j &= u_j{}' \varepsilon_t(u) u_0{}'' \\ u_l &= u_{l-j}{}'', && \text{for } j+1 \leqslant l \leqslant m. \end{aligned}$$

As the equalities above hold in all 2-categories, the rest of the assertion follows. □

**Lemma 6.** Every vertical composition of generalised expansions

$$h_1 \circ \dots \circ h_q : C_1 \dots C_m \longrightarrow D_1 \dots D_m$$

can be rewritten as a normal expansion step

$$o_0 D_1 \ldots o_{m-1} D_m o_m$$

such that

$$h_1 \circ \ldots \circ h_q = o_0 D_1 \ldots o_{m-1} D_m o_m$$

and

$$\overline{h_1} \circ \ldots \circ \overline{h_q} = \overline{o_0}\overline{D_1} \ldots \overline{o_{m-1}}\overline{D_m}\overline{o_m}.$$

Moreover,

$$\overline{h_1 \circ \ldots \circ h_q} = \overline{h}_1 \circ \ldots \circ \overline{h}_q.$$

*Proof.* The proof is similar to the case of generalised contractions. □

**Lemma 7.** If $v_1 \circ \ldots \circ v_n : B_1 \ldots B_m \longrightarrow C_1 \ldots C_r$ is a vertical composition of induced steps, then $r = m$ and there is a normal vertical step $s_1 \ldots s_m$ such that

$$v_1 \circ \ldots \circ v_n = s_1 \ldots s_m \quad \text{and} \quad \overline{v_1} \circ \ldots \circ \overline{v_n} = \overline{s_1} \ldots \overline{s_m}.$$

Moreover,

$$\overline{v_1 \circ \ldots \circ v_n} = \overline{v_1} \circ \ldots \circ \overline{v_n}.$$

*Proof.* First note that the domain and codomain of an induced step are strings of the same length, so $r = m$. Now we proceed by induction on $n$, using the switching operation (0) and the distributive laws (2.3) □

**Lemma 8.** The canonical extension $\overline{(\ )}$ preserves vertical composition.

*Proof.* By Lemmas 2 and 3, each of $g$ and $f$ separately can be written as a vertical composition of single steps, so

$$g \circ f = f_1 \circ \ldots \circ f_n$$

and

$$\overline{g} \circ \overline{f} = \overline{f_1} \circ \ldots \circ \overline{f_n},$$

respectively.

Then, by Lemmas 4, 5, 6 and 7, this vertical composition is equal to

$$f_1 \circ \ldots \circ f_n = o_0 C_1 \ldots C_m o_m \ \circ \ s_1 \ldots s_m \ \circ \ u_0 B_1 \ldots B_m u_m$$

and

$$\overline{f}_1 \circ \ldots \circ \overline{f}_n = \overline{o_0}\,\overline{C_1} \ldots \overline{C_m}\,\overline{o_m} \circ \overline{s_1} \ldots \overline{s_m} \circ \overline{u_0}\,\overline{B_1} \ldots \overline{B_m}\,\overline{u_m},$$

respectively.

By the distributive laws (2.2) and (2.7), we derive

$$f_1 \circ \ldots \circ f_n = o_0 u_0 s_1 \ldots s_m o_m u_m$$

and

$$\overline{f_1} \circ \ldots \circ \overline{f_n} = \overline{o_0}\ \overline{u_0}\,\overline{s_1} \ldots \overline{s_m}\,\overline{o_m}\,\overline{u_m},$$

respectively.

By definition, the canonical extension commutes with horizontal composition, hence

$$\overline{f_1 \circ \ldots \circ f_n} = \overline{f}_1 \circ \ldots \circ \overline{f}_n,$$

and thus

$$\overline{g \circ f} = \overline{g} \circ \overline{f}.$$

This completes the proof of the Theorem in Section 4. □

## Acknowledgments

The authors would like to thank Kosta Došen, who first suggested to them that arrows in a free category could be represented with the help of links. A graphical description of composition is also investigated in Došen and Petrić (2004). The connection between decidability (in lambda-calculus) and normalisation in categories can be traced back to Altenkirch *et al.* (1995) among others. Robert Seely was the first to show that first-order logic has a 2-category structure when formulated with natural deduction (Seely 1979).

## References

Altenkirch, T., Hofman, M. and Streicher, T. (1995) Categorical reconstruction of a reduction-free normalization proof. In: Pitt, D., Rydeheard, D. E. and Johnstone, P. (eds.) Category Theory and Computer Science. *Springer-Verlag Lecture Notes in Computer Science* **953** 182–199.

Barr, M. (1995) Non-symmetric *-autonomous categories. *Theoretical Computer Science* **139** 115–130.

Buszkowski, W. (2002) Cut elimination for the Lambek calculus of adjoints. In: Abrusci, M. and Casadio, C. (eds.) *Papers in formal linguistics and logic*, Bulzoni.

Degeilh, S. and Preller, A. (2005) Efficiency of pregroups and the French noun phrase. *JoLLI.*

Došen, K. (1999) Cut elimination in categories. *Trends in Logic*, Kluwer.

Došen, K. (2002) Simplicial Endomorphisms, Mathematical Institute, SANU (available at `http://arXiv.org/math.GT/0301302`).

Došen, K. and Petrić, Z. (2004) Proof theoretical coherence, Preprint of Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade.

Eilenberg, S. (1972) *Automata, Languages and Machines* **A**, Academic Press.

Eilenberg, S. (1976) *Automata, Languages and Machines* **B**, Academic Press.

Kelly, G. M. and Laplaza, M. I. (1980) Coherence for compact closed categories. *Journal of Pure and Applied Algebra* **19**.

Lambek, J. (1999) Type Grammar revisited. In: Lecomte, A. Lecomte, A., Lamarche, F. and Perrier, G. (eds.) Logical Aspects of Computational Linguistics. *Springer-Verlag Lecture Notes in Artificial Intelligence* **1582** 1–27.

Lambek, J. (2004) Bicategories in algebra and linguistics. In: Ehrhard, T., Girard, J.-Y., Ruet, P. and Scott, P. (eds.) *Linear logic in computer science*, London Mathematical Society Lecture Note Series **316**, Cambridge University Press.

Mac Lane, S. (1971) *Categories for the working mathematician*, Springer Verlag.

Seely, R. (1979) Weak adjointness in proof theory. In: Fourman, M. P., Mulvey, C. J. and Scott, D. S. (eds.) Applications of sheaves. *Springer-Verlag Lecture Notes in Mathematics* **753** 697–710.