

Theory of mind, the three voices and temporality

This paper is a swift gallop across the issues of theory of mind, ending up in what is, for me, the safe fortress of linguistics.

It builds upon a presentation given at the Evolution of Language conference in Leipzig earlier this year.

[BOOKLETS]

There are two booklets accompanying this talk.

The bigger booklet, the Role of Communication Structure, was also given out at Leipzig, and sets out the evolution of language theory.

The smaller booklet, Theory of Mind, sets out **this** talk in more detail.

One of the few stable features across languages is the recognition of three types of being in the universe:

- the self;
- the directly addressable **non**-self;
- and the non-self that is **not** directly addressable.

We refer to these by the conventional terms of the **three persons**, or the **three voices**; and we use a special class of words, pronouns, to represent them.

Different languages use different ranges of pronouns, but they all come down to the three voices that we use to express the actors in a linguistic proposition:

I, you, he-she-it, we, you, they.

The pronouns in a language can be revealing about the interpersonal relationships **culturally** expressed in the language —

both in terms of what is **differentiated** within a voice and what is **not** differentiated.

The Spanish *tu* and *usted* forms are an example of differentiation, and **we** is an example of **non**-differentiation.

In English – and **most** languages – there is no linguistic differentiation between the seven meanings of **we** —

but the gestural signs of BSL are **much more** revealing about who **my group** contains.

This is discussed on **page two** of the smaller booklet.

Because the three voices are so ubiquitous, we should expect them to express something **fundamental** in language communication — possibly something fundamental to **all communication** — And this is precisely what we find.

[DIAGRAM 1 – simple semiotic diagram]

In communication, we see a relationship between:

- the **sender** of a message, who is always the first person, me;
- the **receiver** of the message – always the second person, you;
- and the **referent** of the message – always the third person.

This has been expressed in the very basic standard semiotics model shown.

It also matches with Jakobson's **constitutive factors** of a speech event, and Morris's **basic terms** of semiosis —

Only if, however, they are stripped of linguistic significances.

Non-language communication, in **this** model, does not include Jakobson's **context** or **code**; —

and Morris's **certain conditions**, **interpretants** and **significations** are all part of the **referent**.

This is discussed on **page four** of the smaller booklet.

Non-language signals are completely **context bound** —

the signal **is** the context —

and there is no coding or decoding needed, because the **signal** is also the **code**.

Also, because a **message** is an encoding of a **context**, the **signal** is also the **message**.

In addition, the **voices** are all predetermined.

The **Sender** is always **me** – it does not need to be stated —
the **Receiver** is always **you** – it, too, does not need to be stated —
and the **Referent** is always the **third person**.

However, the referent does not have a communication role **external** to the message —
So the **signal** containing the **message** has to be explicitly about the referent.
Additionally, in non-language communication, the **action** implied by the **signal** is
implicit to both the **referent** and the signal **itself** —
one sign represents both the referent, and the action the **sender** wants the **receiver** to
perform.

[DIAGRAM 2 – merged non-language model and language model]

So, one signal contains one message in one code with one action about one referent.
However, even though all these things are co-identified in **non-language**, they are
still intrinsically present as separable constructs.

In language, things are very different —
Signal, message, code, action and referent are all **separate**.

In language, we are also able to subdivide the message into three parts to create a
simple proposition:
something / does / something.

We traditionally call this the **subject-verb-object** structure of a sentence —
but it reflects a **deeper structure**, which can be identified in **non-language**
communication, too.

Structurally, the components of a linguistic proposition —
Subject-Verb-Object — or **Instigator-Action-Recipient** —
map directly to the components of communication (**Sender-Referent-Receiver**).

However, **semantically** they have become dislocated:

- the **action** invoked by the **referent** is no longer co-identified **with** the referent;

- the **recipient** is no longer always **you**;
- and, most importantly, the **instigator** does not need to be the **sender**.

[DIAGRAM 3 – I leopard / climb-tree you]

When a vervet monkey makes a leopard call, what **happens**, according to this model, is
 [I] [leopard / climb-tree] [you].

It is reflected in the single English word *Leopard!* —

when it is used as a warning.

In **language** we can say *John is warning Mary about the leopard*, separating : —

- the sender, **I** from the Instigator, **John**;
- the receiver, **you**, from the recipient, **Mary**;
- and the referent, **leopard**, from the action, **warning**.

How these separations happen is set out in detail in the **larger** booklet, —

And it shows that the separation of the **Sender** from the **Instigator** is tied up with the ability to produce a model of **ourselves** within our own mind.

Full human language needs a **theory of mind** for **the self** to be able to model situations where the self is **not** the only possible instigator in an activity —
 And this only becomes possible when the **self** is able to model itself.

Intuitively this should be the **first** step in any theory of mind, but in practice it turns out to be the **final** step.

This is a result of the imperatives of evolution —

it is **always** advantageous to be able to predict the actions of others — you can then take steps to counter those actions **before** they have occurred.

However, the cost of predicting your **own** actions is that you have to take a **dispassionate** view of the self —

and how could being dispassionate about yourself be of **advantage**, when all around are **intrinsically passionate** about their own survival and reproduction?

Something **happened** in human evolution that allowed us to answer the question — *what would I do if I were you?*

But it is impossible to even begin **posing** that question in a strictly evolutionary environment.

So what happened is likely to have, at least in part, an explanation that relies on something **other** than direct evolution.

I have some ideas about that, and they are set out in the **conclusion** of the **larger** booklet.

This brings us to the expression of time in language; or, to give it an individual label, temporality.

I have argued that human language is based on the identification of the **self as other**, a level of self-awareness that is likely to be **absent** from other animals.

If the idea has merit, then it should be possible to find evidence of it within language **itself**, and temporality provides an excellent example.

[DIAGRAM 4 – past, present and future with conditionality]

The **traditional** view of time in language is that we can express events on **two** **vectors**.

The **first** extends from the **past** into the **future**, and the **second** extends from **positive certainty**, through a range of probabilities, to **negative certainty**.

The **physical** self, the **self-who-knows**, remains **fixed** in the present — but it can view events as **having occurred** in the past, or as **due to occur** in the future.

It does this by modelling the **self-who-does** into the appropriate timeslot.

This is a powerful capacity, but it is a **mere fragment** of what we do in language.

As humans we are able to divide ourselves into **multiple** instances.

Not only can the **self-who-knows** produce models of the **self-who-does**, —

it can produce models where a **self-who-views** has an **internal model** of the **self-who-does**.

We now have **three** selves, with a **watcher** – watching the **watcher** – watching the **watched**.

In theory, we can insert **more** instantiations of the **self-who-views** between the **self-who-knows** and the self-who-does —

But, in practice, we **seldom** do so.

For instance —

tomorrow at this time I will be about to have finished leaving the country,

or —

three days ago yesterday I had not smoked for a month.

These are **linguistic** examples of **tetrapartite** division of the self.

However, these are more **intellectual exercises** than **regularly used forms**, — and extracting **meaning** from them is difficult.

For **now**, I will concentrate on the **three** selves.

How do the **three selves** work **together** to express **temporality** in language?

The easiest way to demonstrate this is to look at some tenses in English.

Some linguists argue that English has only **two** true tenses, **present** and **past** —

I do and **I did**.

Others argue that **any** expression of temporality on a verb counts as a **tense marker**.

And **others** argue that constructs including auxiliary verbs like **will** and **have** count as tenses —

This last argument is what **I** mean by tense.

In **English**, we have the uncontroversial **past perfect** and **future simple** tenses —

I have done and **I will do**.

In both of these cases the **self-who-knows** is in the **present**, as is the **self-who-views**

—

but the **self-who-does** has been modelled into the **past** or **future**.

The temporal location of the **self-who-does** we can call the **action point**, the **point** at which the **action of doing** takes place.

We can also move the **self-who-views** into the past or future, — and the **self-who-does** then has a **past** or **future** position **relative** to the **self-who-views**.

The location of the **self-who-views** we can refer to as the **viewpoint**.

The **self-who-knows** **cannot** be moved from the present — because **knowing** is the **here-and-now** part of the construct.

The relationship between the selves is **linear** — the **self-who-knows** has a relationship with the **self-who-views**, — and the **self-who-views** has a relationship with the **self-who-does**. But the **self-who-knows** has no direct relationship with the **self-who-does**.

[DIAGRAM 5 – simple tense engine]

We now have a **range** of temporal relationships, — and these are shown on **page eleven** of the booklet.

The tense forms can be **expressed** as follows:

- **I had done.**
- **I was going to do.**
- **I did.**
- **I am doing.**
- **I will do.**
- **I will have done.**
- **I will be going to do.**

This collection of tenses leaves out **several important** components of the linguistic expression of temporality.

These are:

- **continuity or imperfection;**
- **imminence;**

- **conditionality;**
- and **connectivity.**

These components are not **outside** the model, they are **part** of it.

I will describe them **swiftly** to show how they fit in.

First, **Continuity**.

An event can occur in **one** of **three** ways:

- it can be a **single** event, **completed** at the action point —
I shot the sheriff;
- it can be a **single** event **not completed** at the action point —
I'm writing a letter to papa;
- or it can be a **series** of events, **some** of which are complete at the action point, and **some** of which are not —
I go to school every day.

Continuity, therefore, expresses temporality at the **action point**.

Incidentally, some verbs contain continuity **semantically**.

For instance, *I like my sister every day* sounds odd because **like** is a **single** event **not completed** at the action point.

Next, **Imminence**.

An **action point** can occur **close** to a **viewpoint** or further away; —

and a **viewpoint** can be **close** to the **present** — or even **in** the present — or further away.

In **English**, imminence is usually expressed with **adverbials**, like *soon* and *just*.

It can also be indicated by **absolute time** terms, — like *tomorrow* and *last week*.

It is interesting to note that, in **English**, the **adverbials** tend to affect the distance between **viewpoint** and **action point**, —

while the **absolute terms** tend to affect the distance between **viewpoint** and the **present**.

However, this is not an absolute rule —

some adverbials, like *later*, serve in **both** roles.

To summarise, **Imminence**, therefore, expresses the **temporal distances** between the three selves.

Next, **Conditionality**.

This occurs between **now** and the **action point**, —
but **only** if the **viewpoint** is in the **present**.

I may have eaten and *I might eat* are permissible English forms.

I may had eaten is just wrong, —

while *Perhaps I was going to eat*, *perhaps I will have eaten* and *perhaps I will be going to eat* add **uncertainty** but not, strictly, **conditionality**.

Events in the **future** of the viewpoint **already** have the uncertainty that the future is **unknown**, —

and adding an adverbial – *perhaps, maybe, it is likely that* – only **increases** the uncertainty, —

it does not change **certainty** to **conditionality**.

Indeed, the replacement of *will* with *may* illustrates the problem **particularly** well.

I may have eaten does **not** express the same temporality as *I will have eaten*, —

May causes the **viewpoint** to elide into the **present**.

My **current** thinking on this is that conditionality moves the viewpoint through **irreality** and **not** through **time**.

However, this needs more investigation before anything **definitive** can be said.

[DIAGRAM 6 – tense engine with continuity, imminence and conditionality]

This diagram shows where **continuity**, **imminence** and **conditionality** operate in terms of **action point**, **viewpoint** and **now**.

But it **cannot** illustrate how **connectivity** works.

Connectivity is **not** an expression of temporality **within** a construct but **between** constructs.

However, it expresses an **important feature** of the never-ending discourse of language: —

no construct stands alone.

Some connectives carry **low** levels of temporality: —

*I went to the cinema **and** ate popcorn;*

some carry **oblique** levels: —

*I went to the cinema **because** I wanted some popcorn;*

and some carry **high** levels: —

*I went to the cinema **before** I bought the popcorn.*

Connectives allow utterances to place events into a **structured temporal relationship**, —

which is at the **heart** of the human ability to tell stories.

To sum up, **because** we are able to see ourselves as **more** than **one** person, we are able to **overcome** the problem of viewing **everything** from the present.

By modelling ourselves as **future** or **past** entities we create the twin possibilities of **planning** and **reviewing** our actions;

but **this** also raises the possibility of the model of **ourselves** creating a model of **itself**, so that we can **review** our plans **before** they are enacted.

We are a **social** species with **language**;

so, the **complexity** and **functionality** of our mental models is reflected in the **complexities** of **language** —

we can **tell** each other our **plans** and **reviews**.

I hope I have shown that **sense of self** and **self-awareness** are **vital** to language.

They dictate our ability to **express time** within language, and they **even** determine the fact that we have language **at all**.

For **Chomsky**, recursion is — **currently** — the *sine qua non* of language.

But **recursion** is only an **emergent property** of **temporality** —

Which, in turn, is an **emergent property** of **sense of self**.

Exploring **sense of self** is not only more **seminal** and more **productive** than exploring recursion, —
it's **more fun**, too.

Word count: 2450

Max Word count = 2400 for 20 mins