Specifications of a **navigational tool** to help authors to **overcome** the  **problem**.

Michael Zock[1] and Dan Cristea[2,3]

[1] Aix-Marseille Université, CNRS

[2] "Alexandru Ioan Cuza" University of Iași

[3] Institute of Computer Science, Romanian Academy

michael.zock@lif.univ-mrs.fr, dcristea@info.uaic.ro

The beauty, power and magic of **plan B**:

an example, networking

If you want to **meet** the president,

why not try first

to get in **touch** with one of his friends?

or, to change topics

If you are **looking** for a **word** you can't **find**, look for any of its companions (or neighbors):

Outline

1. **The problem** (wordfinding)
2. **Goal :** enhance an existing electronic dictionary to allow for finding quickly and naturally the elusive word.
3. **Analysis of the problem**
   - speech errors, perception, ...
4. **Solutions**
   - **my proposal** (roadmap)

Consider the following
(too often overlooked) **facts**

It is not because something is **stored** that it can readily be **accessed.** This holds definitely for people, but also for machines

- people (amnesia, anomia, TOT, etc.)
- machines

Can you name these objects?

Navigational instrument		sextant
Instrument used in Asia for eating		chopsticks
Hat of a bishop		mitre

My concern

Language **production**

- speaking
- writing

The 3 principal steps

idea

concepts

form

abstract words/lemma
syntactic category
morphology

sound

phonemes
graphemes

The direct way
plan A

The mice are dancing.

Comment

This decomposition (meaning, form, sound) holds not only for texts or sentences, but also for words

The normal situation
a cascaded flow of information

Direct access

Questions

1° How is this **possible** (online processing), i.e. how does our brain manage?

2° Can we achieve something similar via a computer (off-line processing; dictionary consultation)?

- speed
- accuracy
- success in wordfinding

Questions

3° Why do we have problems?

4° Can we draw on the **mental lexicon** to improve the **electronic dictionaries** of tomorrow?

- If not, why so?
- If yes, on what specific aspects

Concerning storage or acces words are just like any other object

- **What** can we observe when we **fail** to remember someone's name?
- People (nearly) always remember '**something**' concerning the target word.
- Note that finding peoples' names is just a special kind of access problem. Whether you look for *words*, peoples' *names* or any *object* (in your household or else) is but a matter of indexing, i.e. organization.

Example : name of a person

Name of **actor**

Film: Silence of the lambs

Role : Hannibal Lecter

Name actor : ???

First name : Anthony

Look for actors whose first name is ‘Anthony’

Anthony Quinn?
Perkins?
Hopkins?

Comment

We always **remember something** concerning the person whose name is eluding us

- age
- place of meeting
- event

.

Work on the TOT-phenomenon revealing what people know

Information concerning the form of the target word

a) number of syllables
- first and last syllable (bathtub effet)

b) grammatical information
- part of speech
- gender
- colloquial expression

c) origine (eg. Greek, latin)

d) target word: when presented a list containing the target word they will recognize it immediately and unmistakingly.

Information concerning the meaning of the target word

Parts of the meaning

- mocha: coffee beverage flavored with milk, sugar, and cocoa

Relations to other concepts or words (associations)

- Mocha : town and port in southern Yemen at the red sea
- Starbucks: place where this beverage is served

Since the user knows a lot of all this

Let's use it, and start from there.

Question : how?

Access should be based on what?

1. form (rhymes : write right; prefix/suffix,...)
2. meanings (or meaning elements) of the target word
3. concepts or words related to the target word
   - ➠ lexical relations (synonyme, antonym, hypo/hypernym,…)
   - ➠ associations

Why do we have access problems?

Fundamental differences between **language** (forms) and **thought** (concepts)

- arbitrary link between the two (horse-cheval)
- underspecification
- interference: on the vertical (synonyms) and horizontal axis (co-occurences)
- distribution: the three main components of a word (*meaning-form-sound*) are stored separately

Arbitrary link

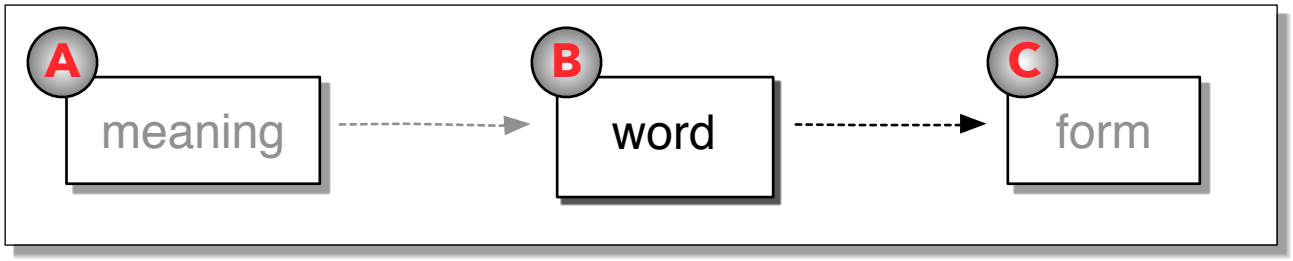
mouse

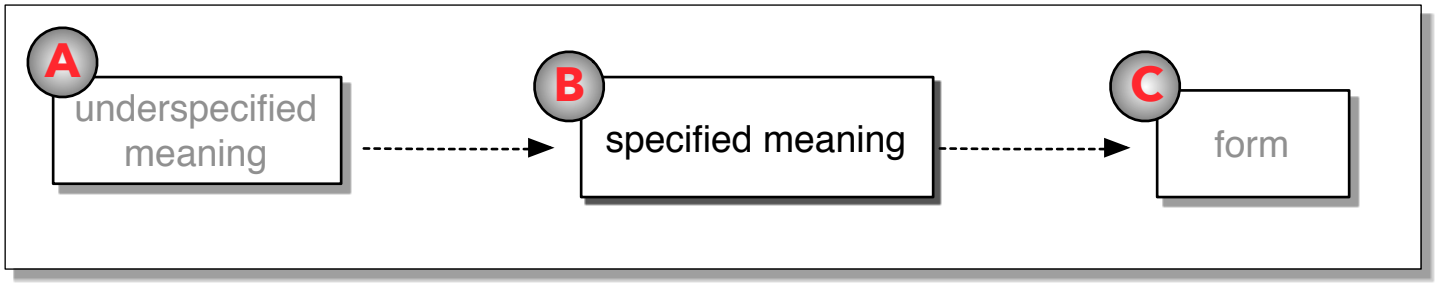
Important note:

- Words and concepts are fundamentally different.
- We (generally) don't think in terms of words, but rather in terms of concepts.
- If we thought in terms of words we would never experience a word-access problems, we would just use the string representing both our ideas (concepts) and words. Yet, this is not quite what we observe in natural settings (spontaneous speech).

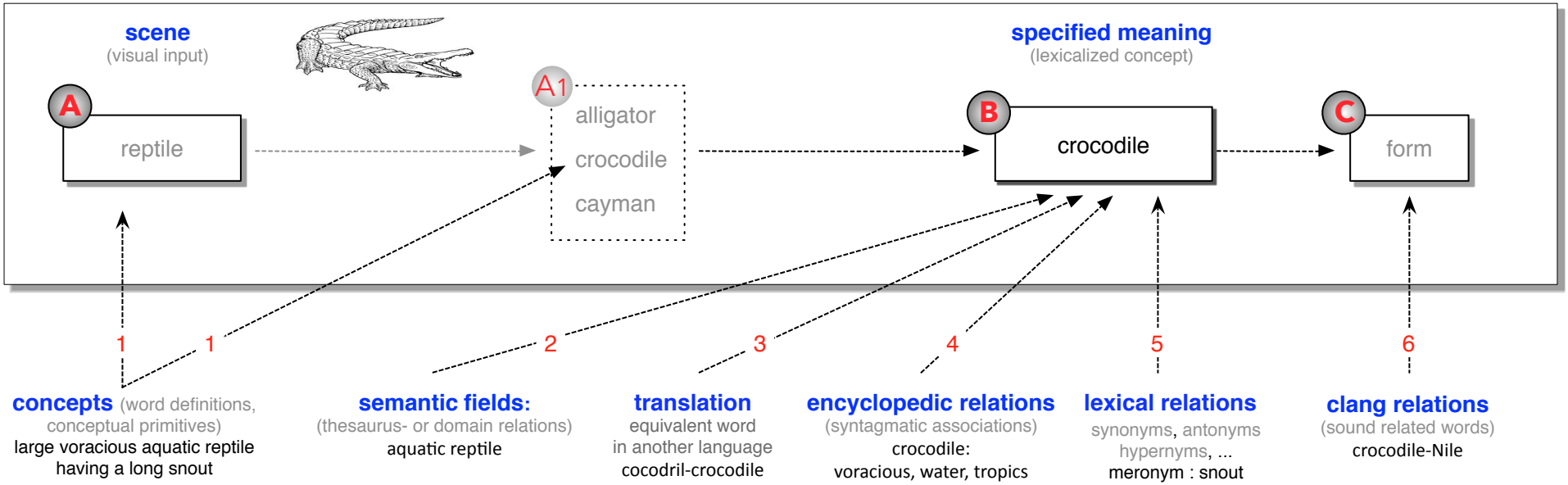
Underspecification

From mind to mouth:
the progressive synthesis of what
we tend to call a word



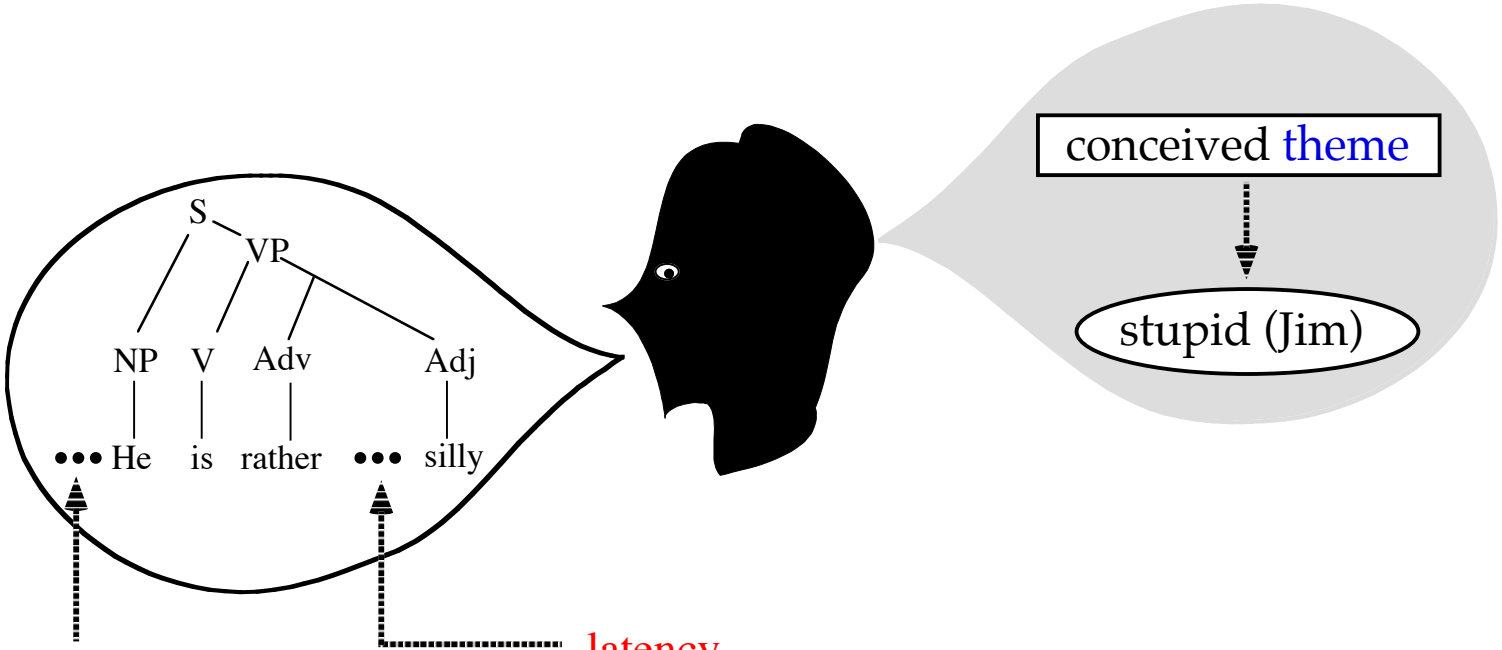


From mind to mouth:
the progressive synthesis of what most of us call a word



Interference : telescoping of competing elements

What do you think of Jim ?

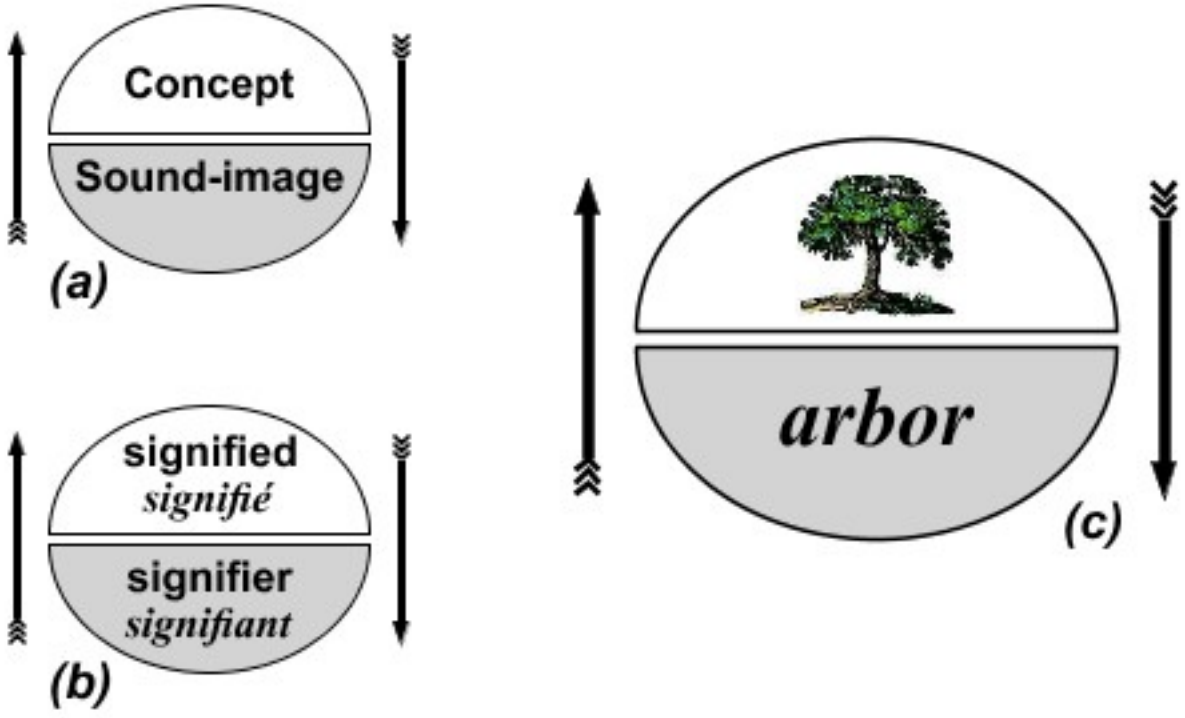


latency
to conceive
the theme :
`stupid (Jim)`

latency
to weigh between
"stupid" and "silly"

He is rather stilly.

Distribution and the illusion of words as holistic entities



Saussure's conception of the 'sign'

Access vs. activation

(continued)

"A potentially counterintuitive idea is that the individual sounds of words are **assembled anew** each time they are spoken rather than **retrieved** as **intact wholes**. Yet, patterns of speech errors and latency data suggest that this is the case. "

Zenzi M. Griffin and Victor S. Ferreira,
Properties of Spoken Language Production, page 35.

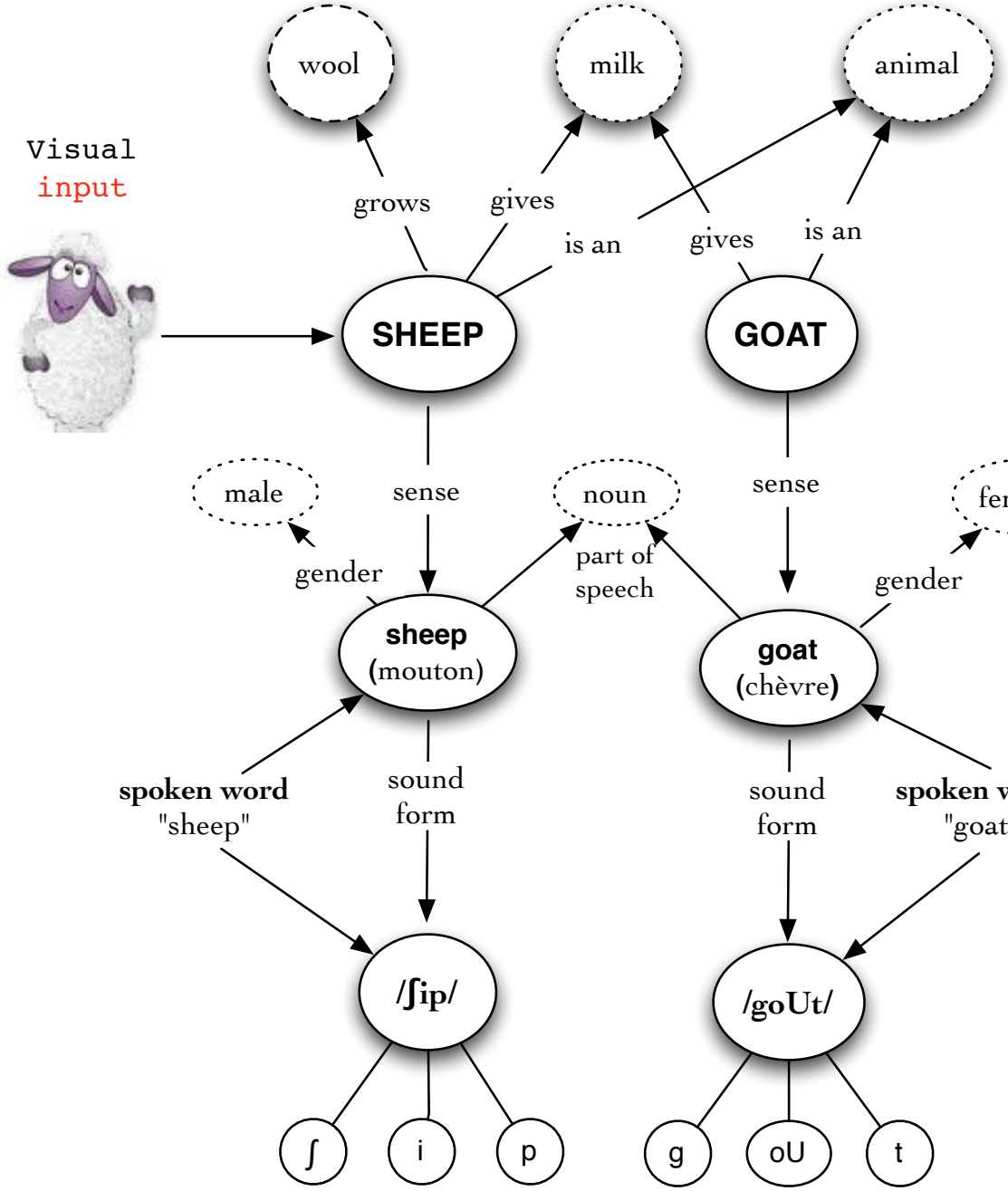
In **Handbook of Psycholinguistics**
Traxler, M. and Gernsbacher, M. A. (Eds.), 2006

Evidence

1. TOT (we do know **fragments** of the word)
2. Speech **errors** at the **different levels**

   - semantics : take the first to the left (target: right)
   - syntax : I make the kettle on (targets: make some tea + put the kettle on)
   - morphology : slicely thinned (target: thinly sliced)
   - sound/phonology : histerical (target: historical)

Levelt's model



Why do we have word-access problems, or, what happens when we are in this state?

1. Competition between the elements at the various levels
2. Similarity between certain elements, hence potential danger of interference and telescoping of information
3. Activation is gradual and relative rather than absolute (all or nothing). For example, we say: it's on the tip of my tongue

Idea (intention of communication) - expression

Idea :

request

```
(make drawing_of
make drawing, you
make drawing, for me)
```

Expression : Will you draw me a sheep!

The problem of finding the (rootform) of words

Input	**Meaning**
Will you draw me a 	woolly usually horned ruminant mammal related to the goat
Semantic candidates	mutton, ram, ewe, lamb, sheep, goat, bovid, ovis
Phonological candidates	cheap, jeep, schliep, seep, sheep, sleep, steep, streep, sweep
Output	/ʃiːp/ - sheep

Suggestion

If plan A doesn't work, use plan B

- plan A (direct route)
- plan B (indirect route)

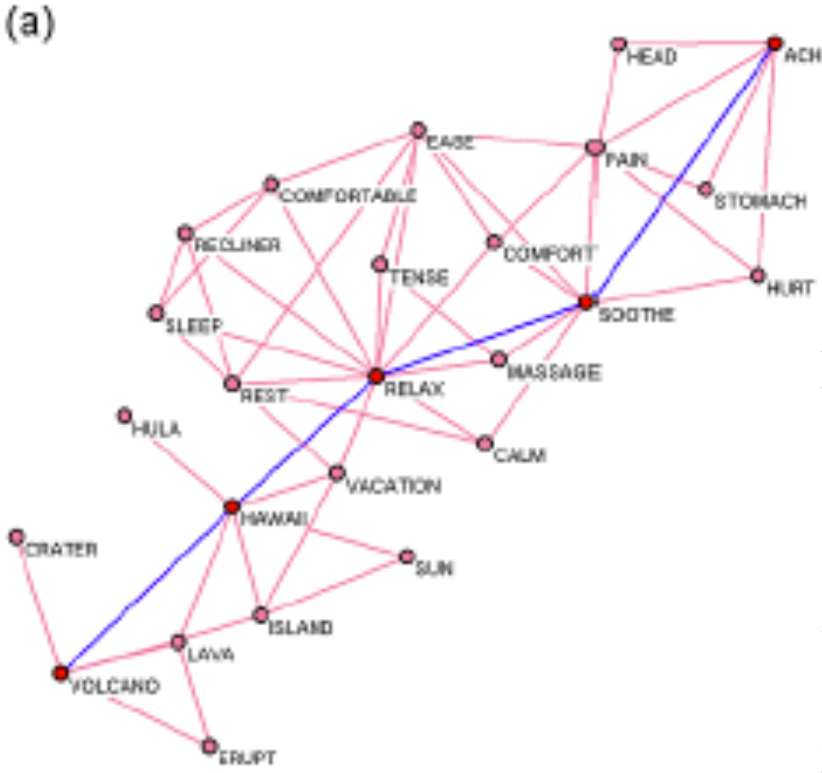
How to overcome the TOT-problem?

Link words and build an **index**

Question: **what kind** of?

Answer : use **associations**

Navigation in an associative network



Since **search** takes will take place within a semantic network, i.e. a graph where all words (nodes) are related (via a certain kind of association), search consists in entering this network at any point and follow the links to get from the starting point (source word, SW) to the end (target word, TW). This latter may be directly related to the initial input, i.e. SW (direct association/neighbour; distance 1) or not (indirect association).

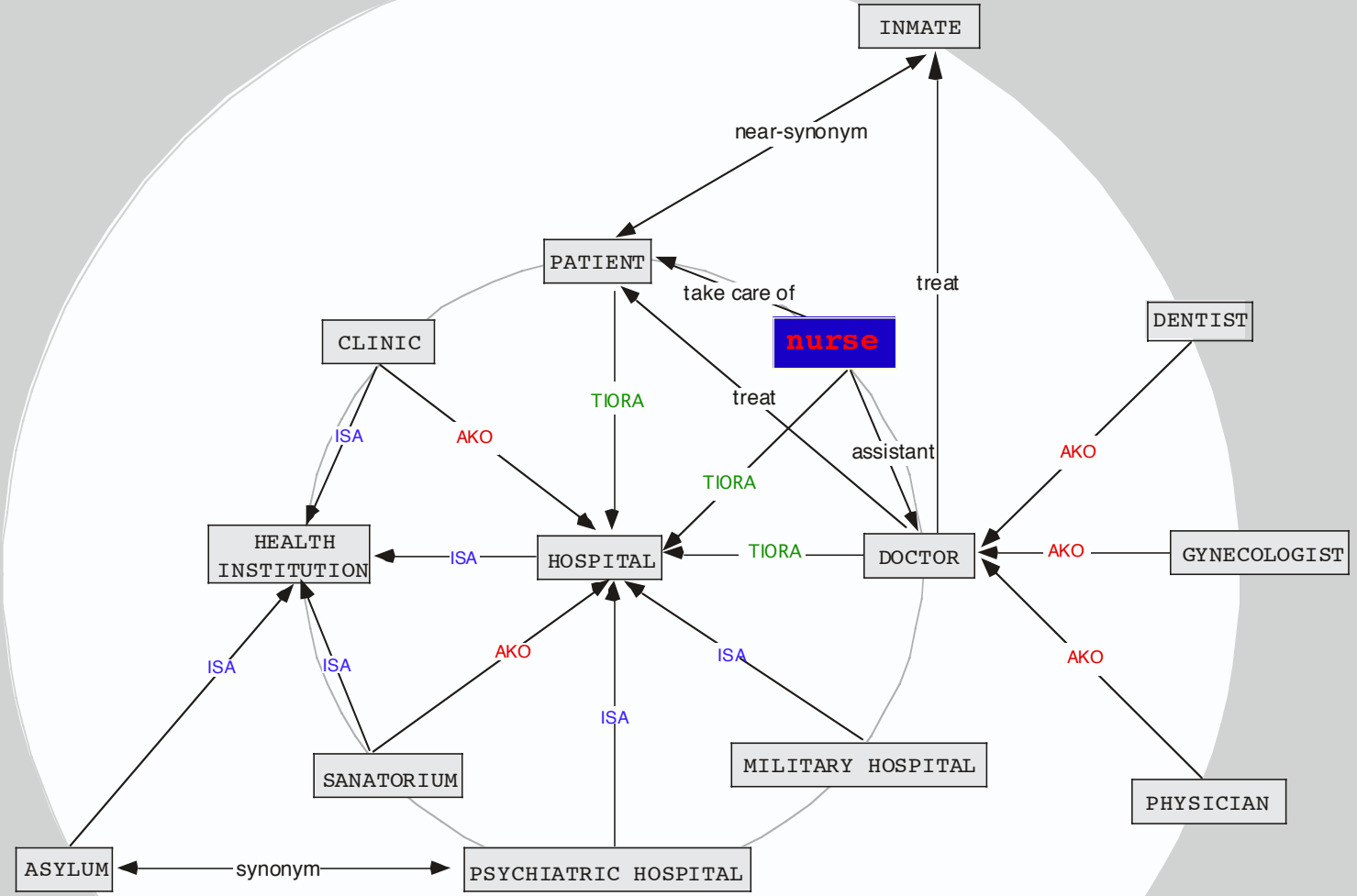
Note that the user **knows** the starting point, but **not** the end-point (target).

Let's put this to work and take an example

word you are looking for (**target word**) ➡ **nurse**

word coming to your mind (**source word**) ➡ **hospital**

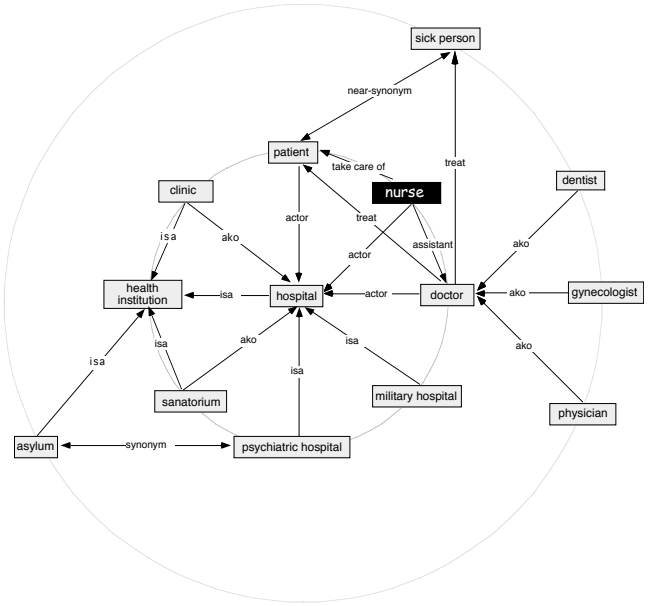
Internal Representation



Links must be interpretable to allow for navigation

Show only what's useful

Internal representation



AKO

- clinic
- sanatorium

ISA

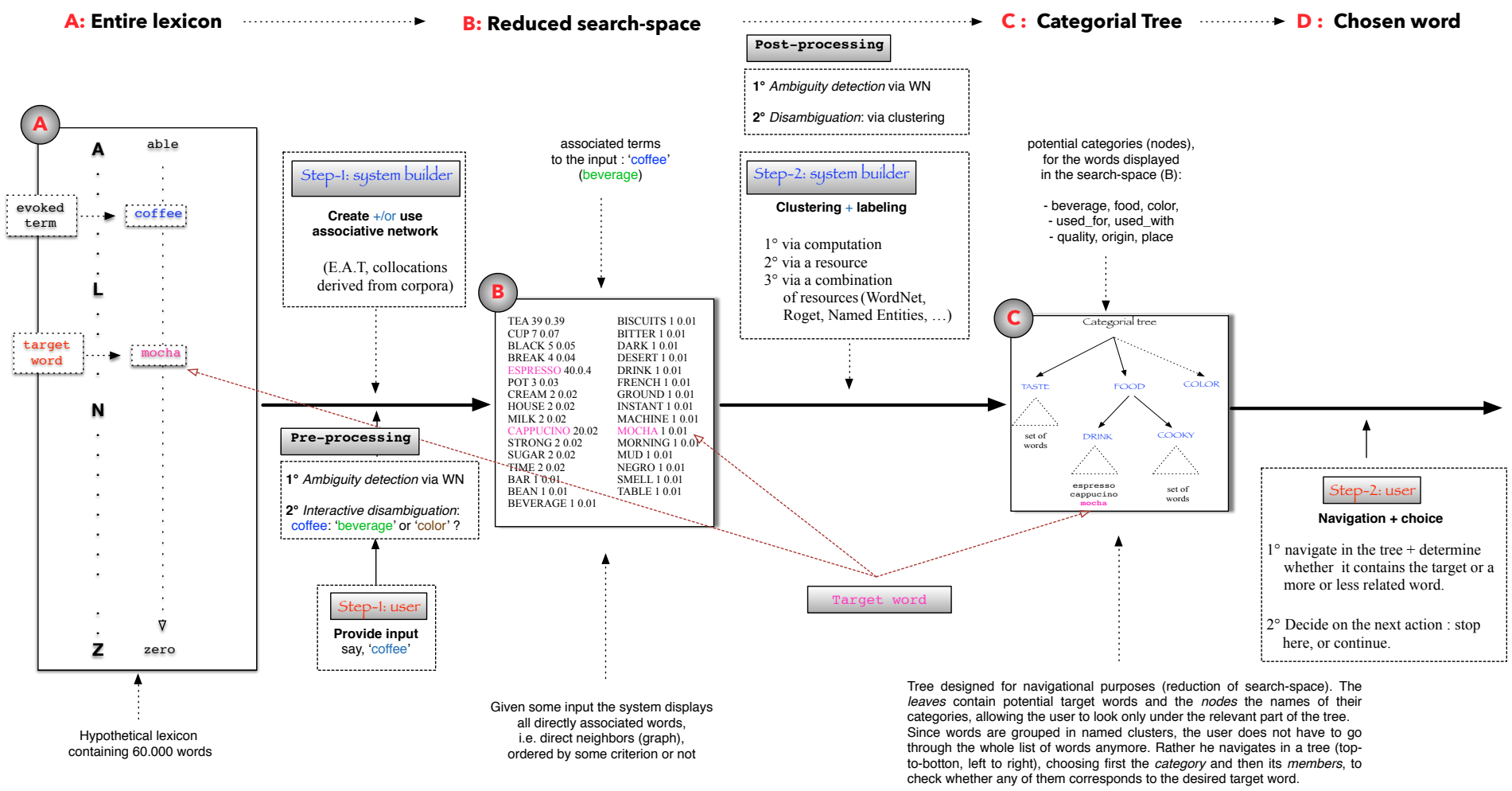
- military hospital
- psychiatric hospital

ACTOR

- doctor
- patient
- nurse

The **nature** of the **problem** of search,
the **framework** of our approach
and its **solution** in a nutshell

How to access the word stuck on the tip of your tongue?



Conclusion

I have presented here some ideas of how to build a resource likely to help authors to overcome the TOT-problem.

I have strongly pleaded for the potential of word associations. While one can certainly rely on the words composing the definition of the target word (meaning, plan A, the normal route), a lot more can be done by using word associations (plan B).

Conclusion

Of course, a lot more work is needed. In particular, we need to

- get the right resources or corpora
- extract the links
- name them and
- build the application allowing to perform the here-described search
- evaluate the tool

Thanks for hanging in!

Dan will tell you now
how to get
all this to work !