Conceptual vectors for NLP

MMA 2001

Mathieu Lafourcade
LIRMM - France
www.lirmm.fr/~lafourca

Objectives

- Semantic Analysis
  - Word Sense Disambiguation
  - Text Indexing in IR
  - Lexical Transfer in MT
- Conceptual vector
  - Reminiscent of Vector Models (Salton and all.) & Sowa
  - Applied on preselected concepts (not terms)
  - Concepts are not independent
- Propagation
  - on morpho-syntactic tree (no surface analysis)

Conceptual vectors

- An idea

  = a combination of concepts = a vector

- The Idea space

  = vector space

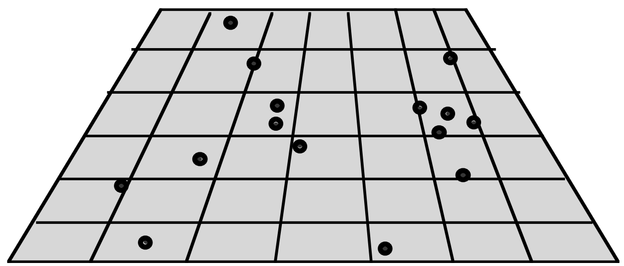
- A concept

  = an idea = a vector

  = combination of itself + neighborhood

- Sense space

  = vector space + vector set



Conceptual vectors

- Annotations
  - Helps building vectors
  - Can take the form of vectors

- Set of k basic concepts — example
  - Thesaurus Larousse = 873 concepts
  - A vector = a 873 uple
  - Encoding for each dimension $C = 2^{15}$

Conceptual vectors

- Example : cat
  - Kernel

    c:mammal, c:stroke

    <… mammal … stroke …>

    <… 0,8 … 0,8 … >
  - Augmented

    c:mammal, c:stroke, c:zoology, c:love …

    <… zoology … mammal… stroke … love …>

    <… 0,5 … 0,75 … 0,75 … 0,5 … >
  - + Iteration for neighborhood augmentation

    Finer vectors

Vector space

- Basic concepts

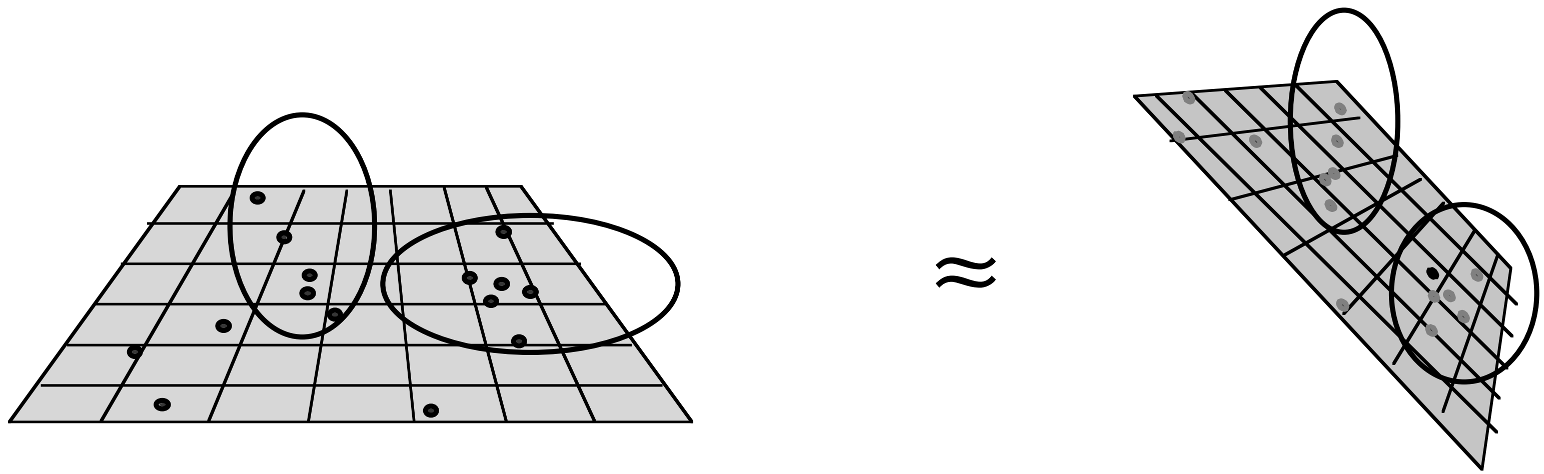
  are not independent

- Sense space

  = Generator Space of a real k’ vector space (unknown)

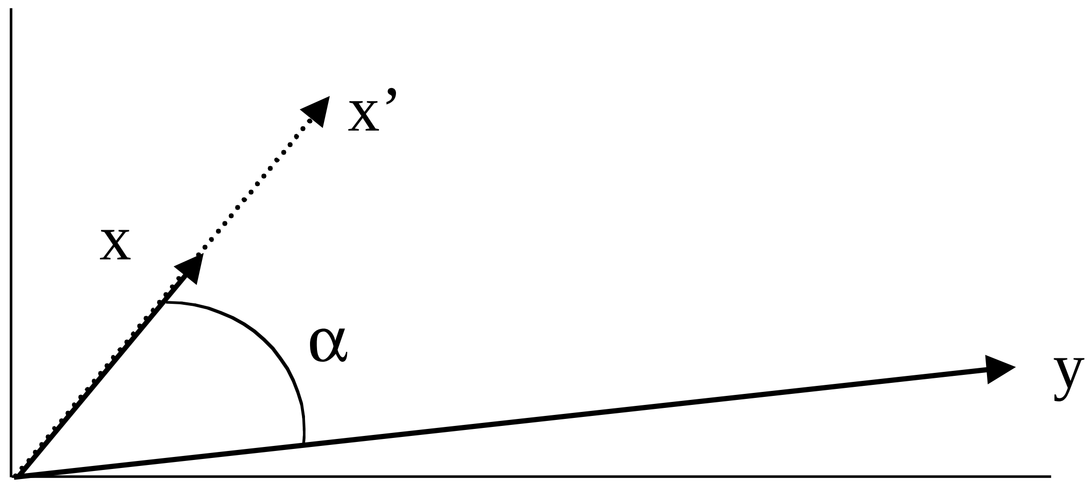
  = Dim k’ <= k

- Relative position of points



Conceptual vector distance

- Angular Distance $D_A(x, y)$ = angle (x, y)
  - $0 <= D_A(x, y) <= \pi$
  - if 0 then colinear - same idea
  - if $\pi/2$ then nothing in common
  - if π then $D_A(x, -x)$ with -x as *anti-idea* of x



Conceptual vector distance

- Distance = acos(similarity)

$$D_A(x, y) = \mathrm{acos}(\sqrt{((x_1 y_1 + \ldots + x_n y_n) / |x| |y|)})$$

$D_A(x, x) = 0$

$D_A(x, y) = D_A(y, x)$

$D_A(x, y) + D_A(y, z) \geq D_A(x, z)$

$D_A(0, 0) = 0$ and $D_A(x, 0) = \pi / 2$ by definition

$D_A(\alpha x, \beta y) = D_A(x, y)$ with $\alpha . \beta > 0$

$D_A(\alpha x, \beta y) = \pi - D_A(x, y)$ with $\alpha . \beta < 0$

$D_A(x+x, x+y) = D_A(x, x+y) \leq D_A(x, y)$

Conceptual vector distance

- Example
  - D_A(tit, tit) = 0
  - D_A(tit, passerine) = 0.4
  - D_A(tit, bird) = 0.7
  - D_A(tit, train) = 1.14

  - D_A(tit, insect) = 0.62

tit = kind of insectivorous passerine …

Conceptual lexicon

- Set of (word, vector) = (w, ν)*

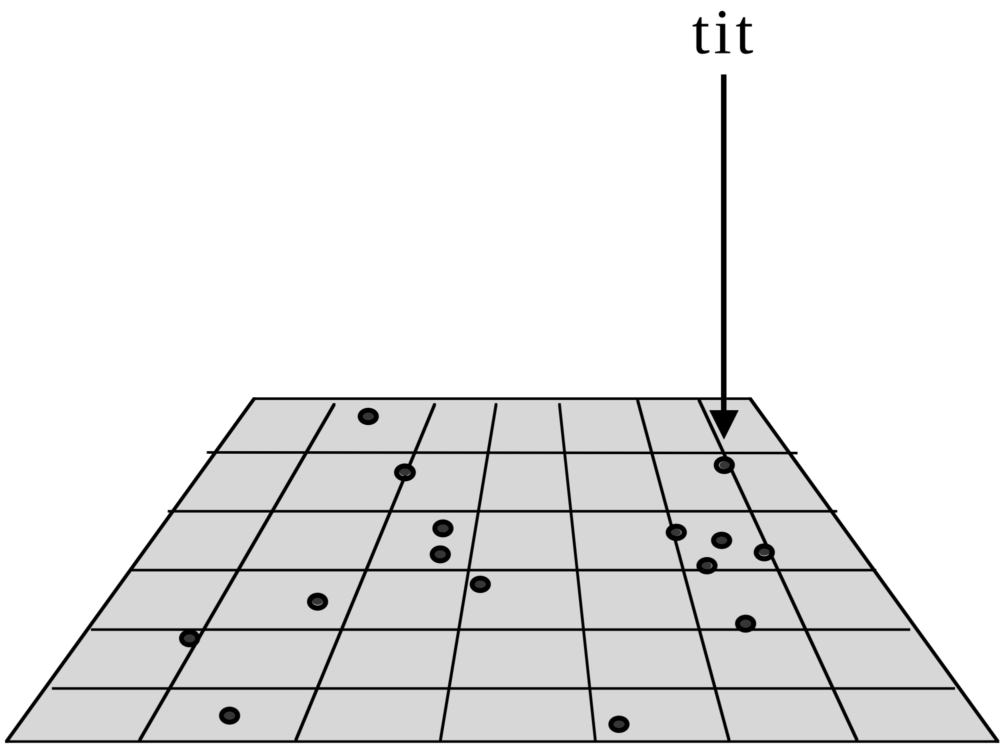
- Monosemy

  word

  → 1 meaning

  → 1 vector

(w, ν)



Conceptual lexicon

Polyseme building

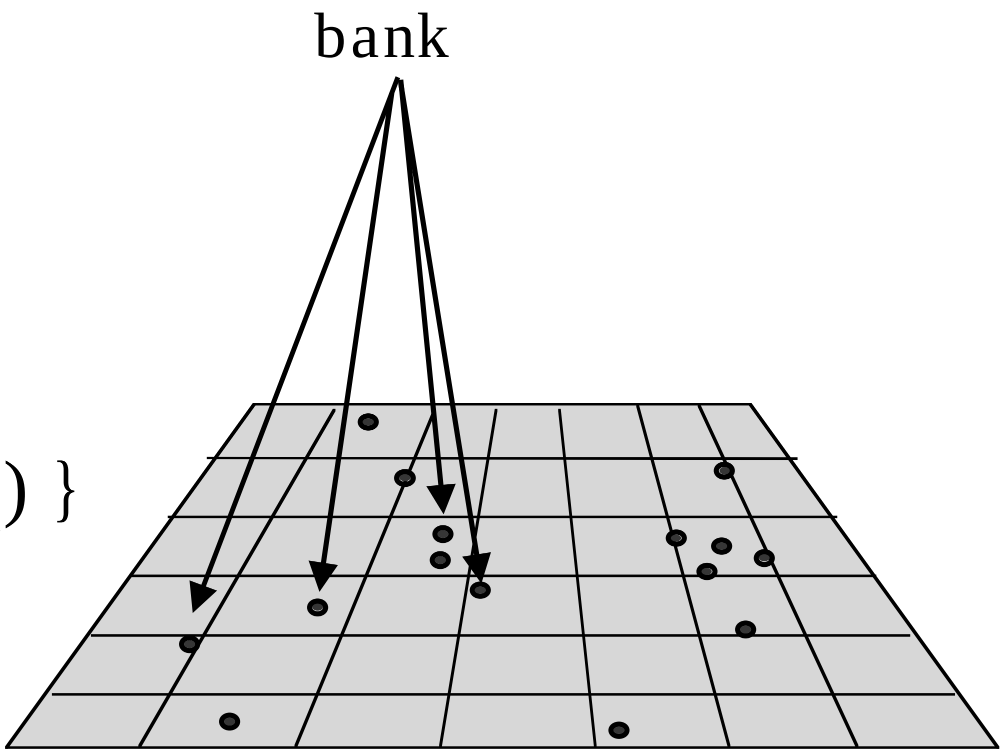
- Polysemy

  word

  → n meanings

  → n vectors

$\{(w, \nu),$
$(w.1, \nu_1) \ldots (w.n, \nu_n)\}$

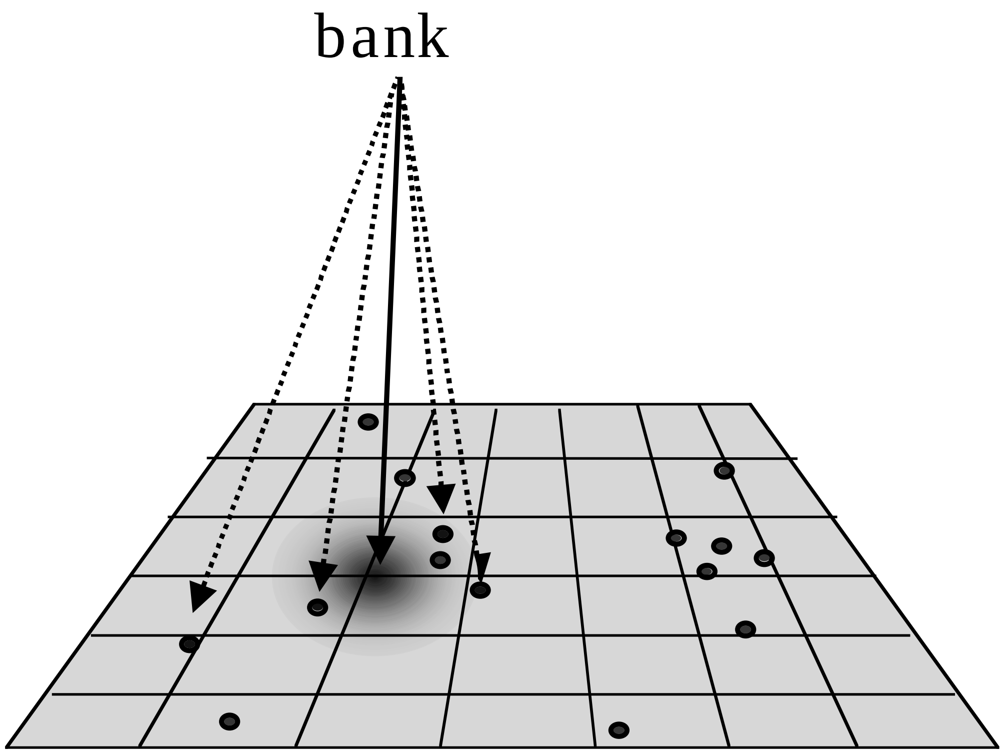


Conceptual lexicon

Polyseme building

- $\nu(w) = \sum \nu(w.i) \quad = \sum \nu.i$

- *bank =*
  - *1. Mound,*
  - *3. river border, …*
  - *2. Money institution*
  - *3. Organ keyboard*
  - *4. …*

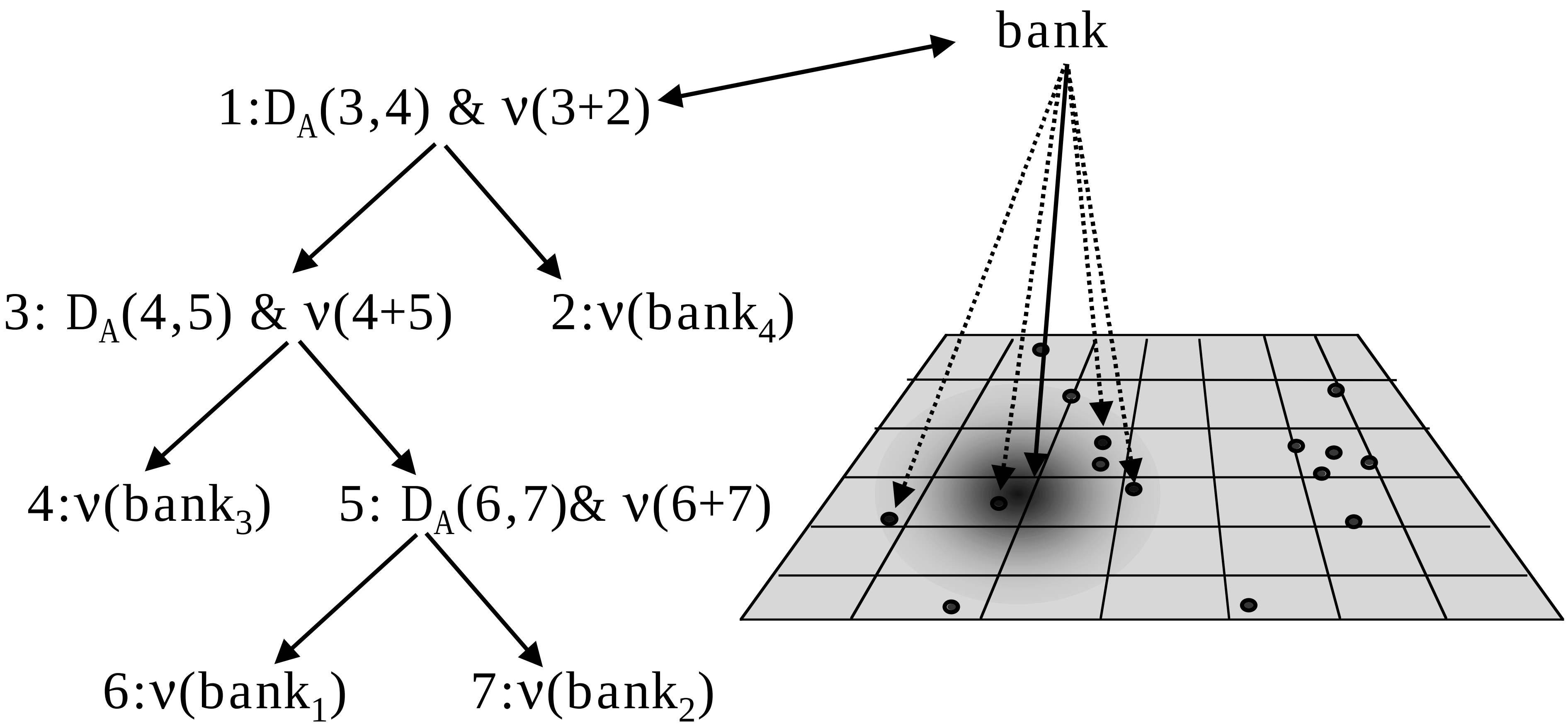


Squash isolated meanings against numerous close meanings

Conceptual lexicon

Polyseme building

- ν(w) = classification(w.i)



Lexical scope

- $LS(w) = LS_t(\tau(w))$

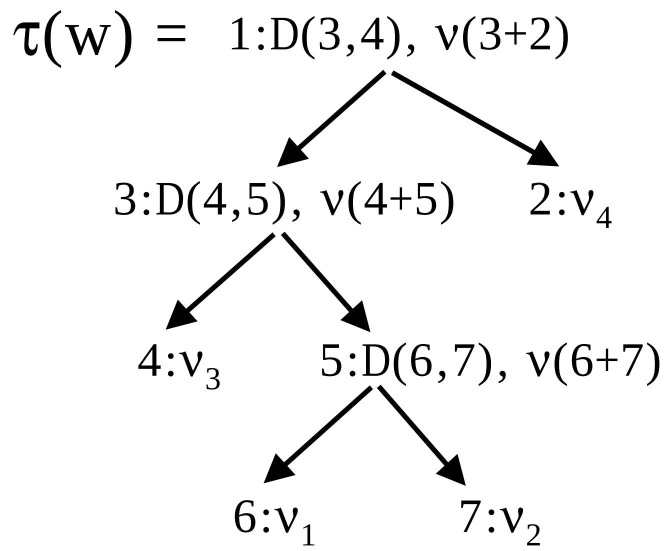
$$LS_t(\tau(w)) = 1 \qquad \textit{if } \tau \textit{ is a leaf}$$

$$LS_t(\tau(w)) = (LS(\tau_1) + LS(\tau_2)) / (2-\sin(D(\tau(w)))) \qquad \textit{otherwise}$$

- $\nu(w) = \nu_t(\tau(w))$

$$\nu_t(\tau(w)) = \nu(w) \qquad \textit{if } \tau \textit{ is a leaf}$$

$$\nu_t(\tau(w)) = LS(\tau_1)\nu_t(\tau_1) + LS(\tau_2)\nu_t(\tau_2) \qquad \textit{otherwise}$$



Can handle duplicated definitions

Vector Statistics

- Norm (N)
  - [0 , 1] * C (2^{15}=32768)
- Intensity (I)
  - Norm / C
  - Usually I = 1
- Standard deviation (SD)
  - SD^2 = variance
  - variance = $1/n * \sum(x_i - \mu)^2$ with μ as the arith. mean

Vector Statistics

- Variation coefficient (CV)

$$CV = SD / mean$$

No unity - Norm independent

Pseudo Conceptual strength

If A Hyperonym B $\Rightarrow$ CV(A) > CV(B)

(we don't have $\Leftarrow$)

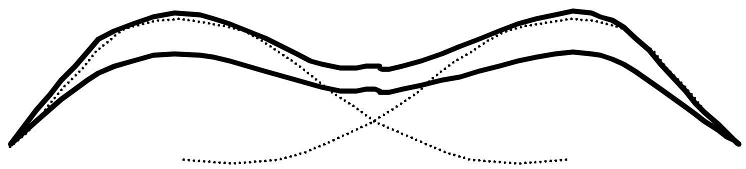
- vector « fruit juice » (N)
  → MEAN = 527, SD = 973
- vector « drink » (N)
  → MEAN = 443, SD = 1014

CV = 1.88

CV = 2.28

Vector operations

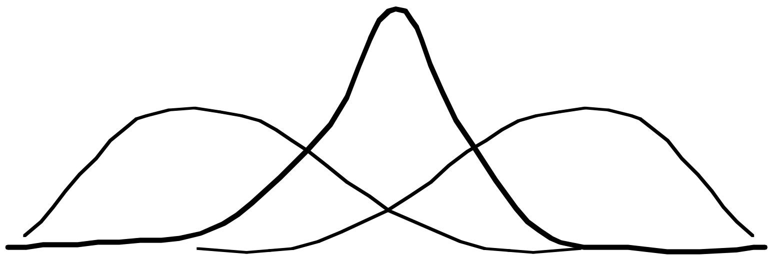
- Sum
  - $V = X + Y \Rightarrow v_i = x_i + y_i$
  - Neutral element : 0
  - Generalized to *n* terms : $V = \sum V_i$
  - Normalization of sum : $v_i \ / |V| * c$



Kind of mean

Vector operations

- Term to term product
  - $V = X \otimes Y \Rightarrow v_i = x_i * y_i$
  - Neutral element : 1
  - Generalized to *n* terms $V = \prod V_i$



Kind of intersection

Vector operations

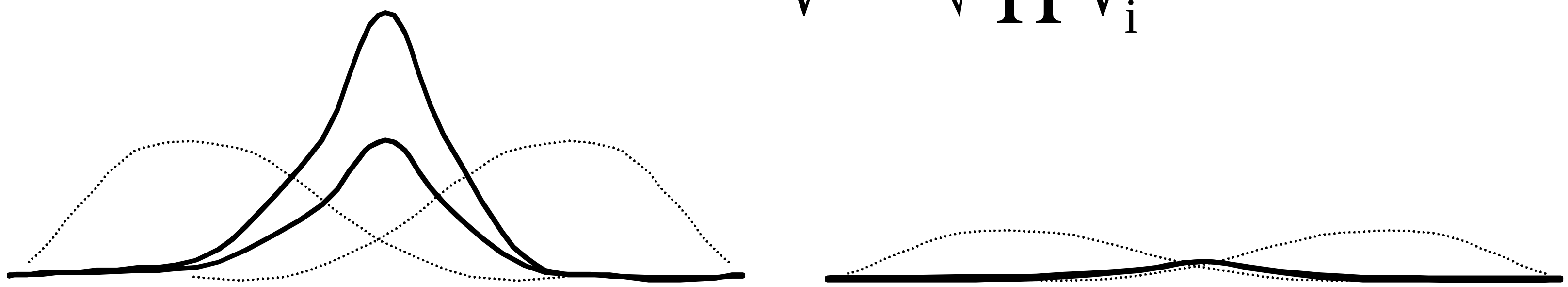
- Amplification
  - $V = X \wedge n \Rightarrow v_i = sg(v_i) * |v_i| \wedge n$

    $\sqrt{V} = V \wedge 1/2$ and $\sqrt[n]{V} = V \wedge 1/n$

    $V \otimes V = V \wedge 2$ *if* $\forall\, v_i \geq 0$

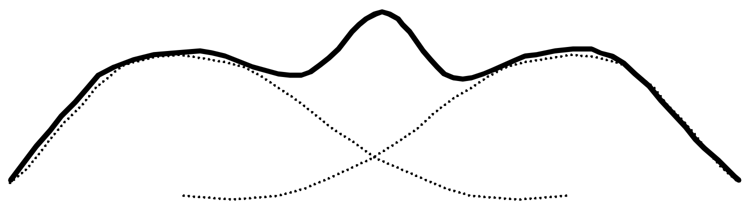
  - Normalization of *ttm* product to *n* terms

$$V = \sqrt[n]{\prod V_i}$$



Vector operations

- Product + sum
  - $V = X \underline{\otimes} Y = (X \otimes Y) + X + Y$
  - Generalized *n* terms : $V = \sqrt[n]{\prod V_i} + \sum V_i$
  - Simplest request vector computation in IR



Vector operations

- Subtraction
  - $V = X - Y \Rightarrow v_i = x_i - y_i$
- Dot subtraction
  - $V = X \dot{-} Y \Rightarrow v_i = \max(x_i - y_i, 0)$
- Complementary
  - $V = C(X) \Rightarrow v_i = (1 - x_i/c) * c$
- etc.

Intensity Distance

- Intensity of normalized *ttm* product
  - $0 \leq I(\sqrt{}(X \otimes Y)) \leq 1 \quad if\ |x| = |y| = 1$

$$D_I(X, Y) = \mathrm{acos}(I(\sqrt{}\, X \otimes Y))$$

  - $D_I(X, X) = 0$ and $D_I(X, 0) = \pi/2$

D_I(tit, tit)	= 0	(D_A = 0)
D_I(tit, passerine)	= 0.25	(D_A = 0.4)
D_I(tit, bird)	= 0.58	(D_A = 0.7)
D_I(tit, train)	= 0.89	(D_A = 1.14)
D_I(tit, insect)	= 0.50	(D_A = 0.62)

Relative synonymy

- $Syn_R(A, B, C)$ — C as reference feature

$$Syn_R(A, B, C) = D_A(A \otimes C, B \otimes C)$$

D_A(coal,night)	= 0.9
Syn_R(coal, night, color)	= 0.4
Syn_R(coal, night, black)	= 0.35

Relative synonymy

- $Syn_R(A, B, C) = Syn_R(B, A, C)$
- $Syn_R(A, A, C) = D(A \otimes C, A \otimes C) = 0$

- $Syn_R(A, B, 0) = D(0, 0) = 0$
- $Syn_R(A, 0, C) = \pi/2$

- $Syn_A(A, B) = Syn_R(A, B, 1)$

  $= D(A \otimes 1, B \otimes 1)$

  $= D(A, B)$

Subjective synonymy

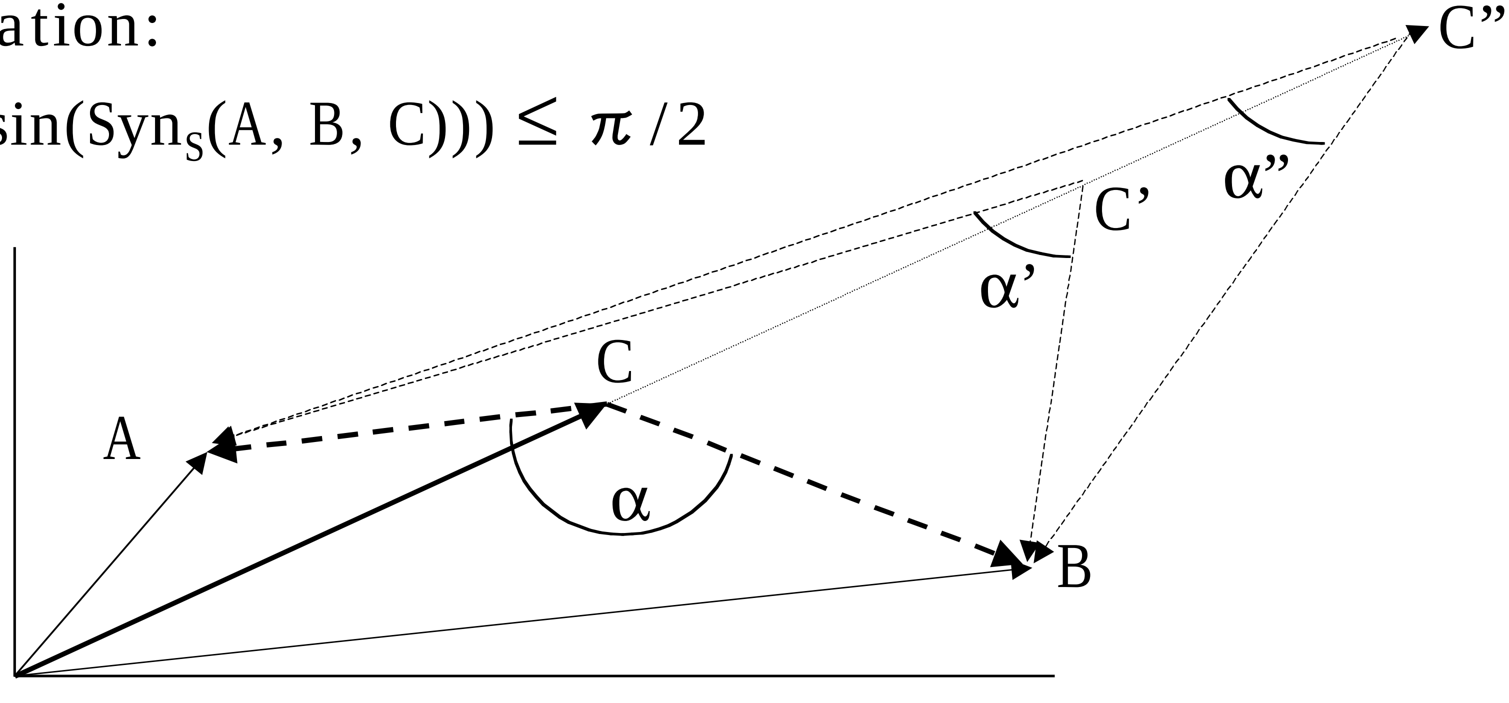
- $Syn_S(A, B, C)$ — C as point of view

$$Syn_S(A, B, C) = D(C-A, C-B)$$

$0 \leq Syn_S(A, B, C) \leq \pi$

normalization:

$0 \leq \mathrm{asin}(\sin(Syn_S(A, B, C))) \leq \pi/2$



Subjective synonymy

When $|C| \to \infty$ then $Syn_S(A, B, C) \to 0$

$Syn_S(A, B, 0) = D(-B, -A) = D(A, B)$

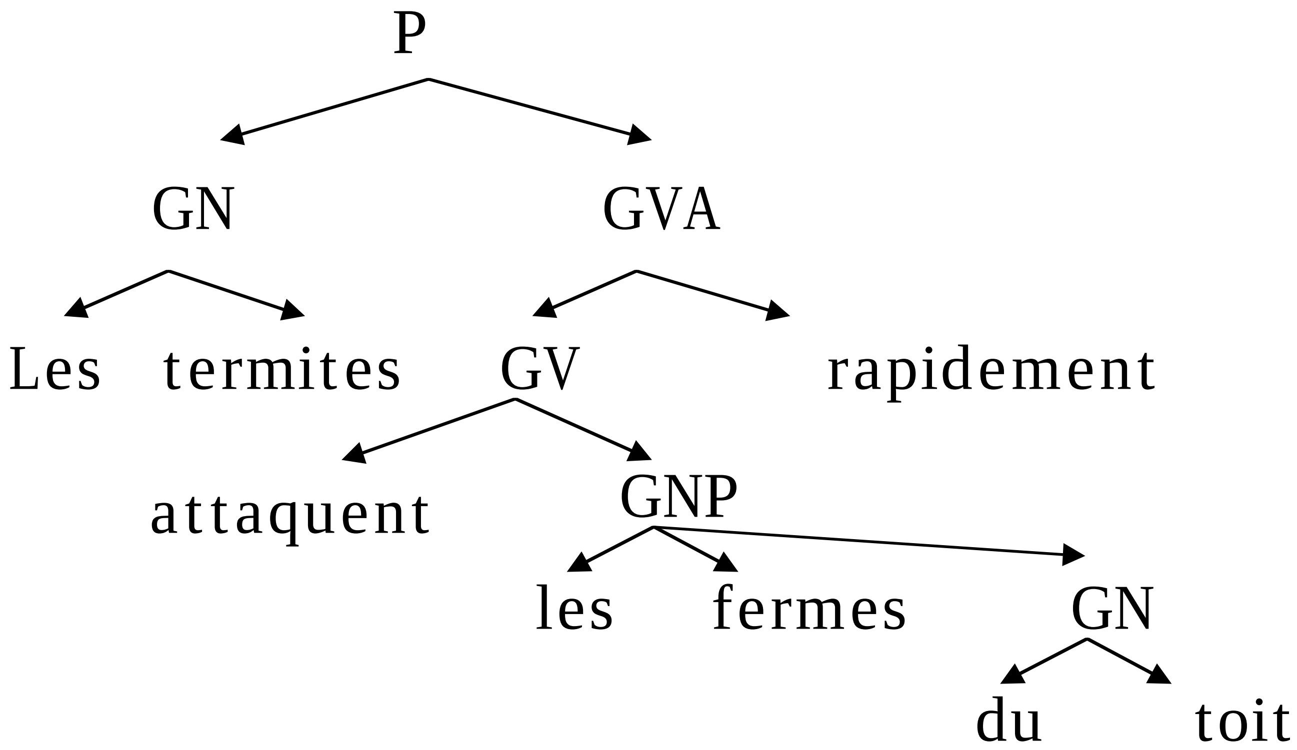
$Syn_S(A, A, C) = D(C-A, C-A) = 0$

$Syn_S(A, B, B) = Syn_S(A, B, A) = 0$

- Syn_S(tit, swallow, animal) = 0.3
- Syn_S(tit, swallow, bird) = 0.4
- Syn_S(tit, swallow, passerine) = 1

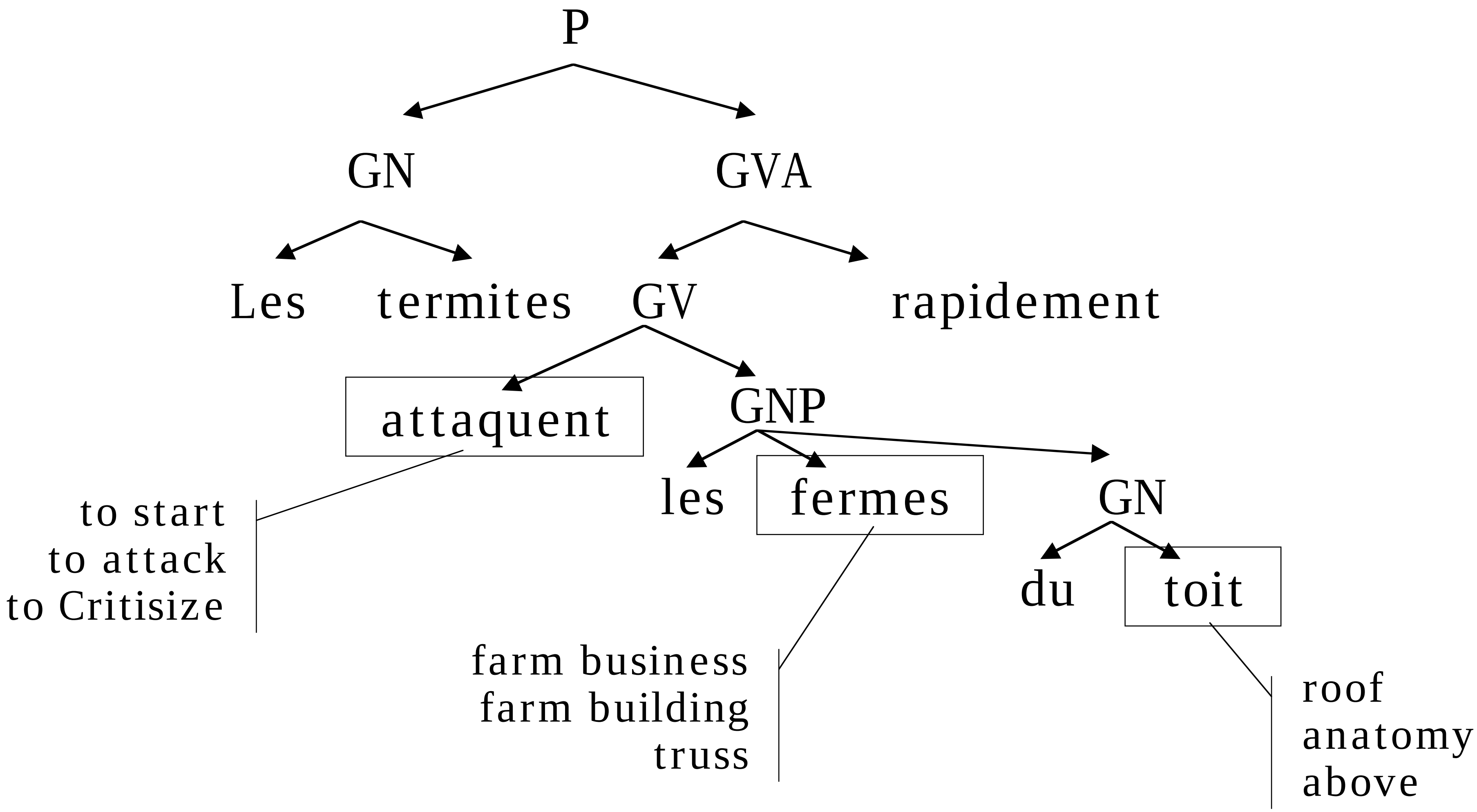
Semantic analysis

- Vectors propagate on syntactic tree



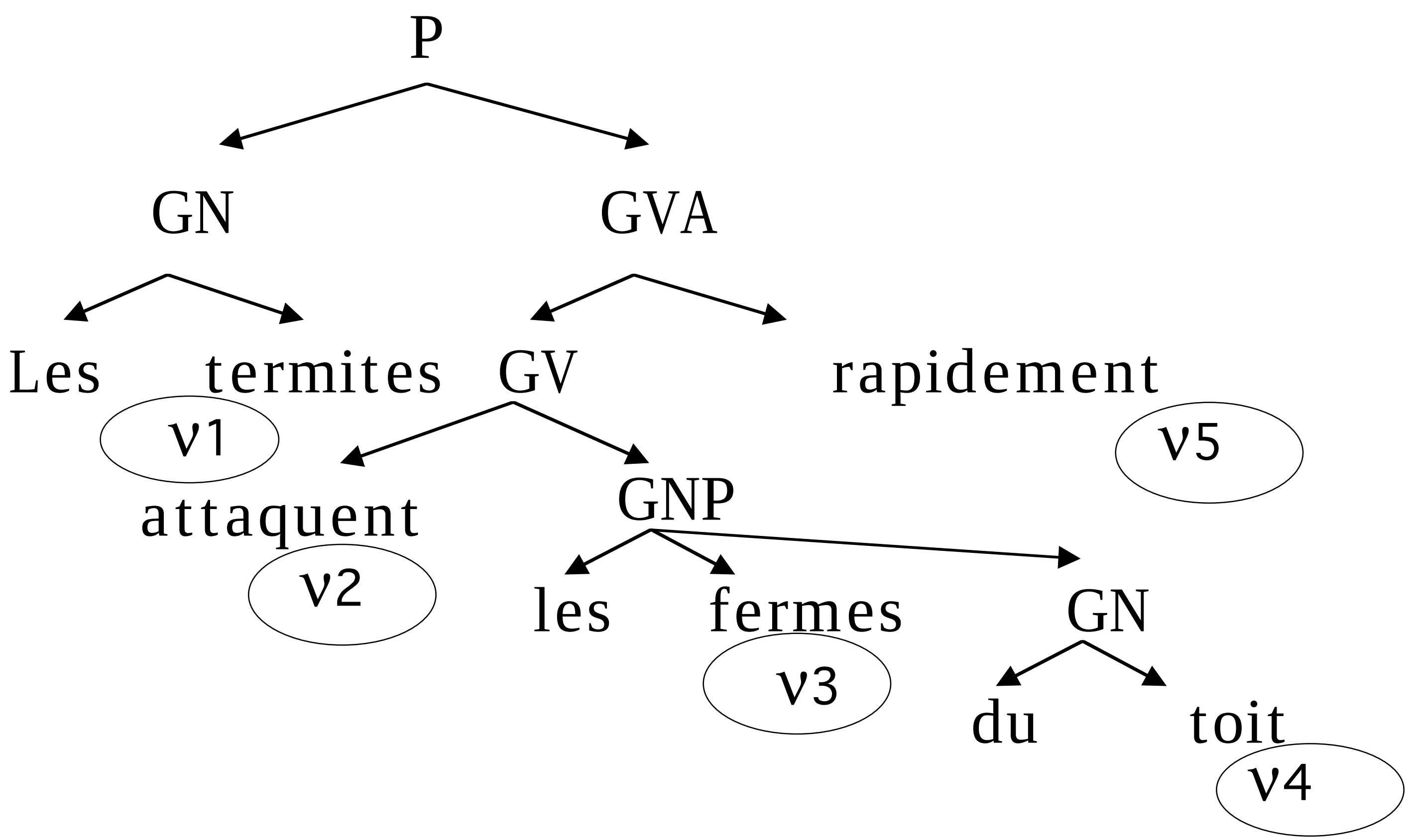
The white ants strike rapidly the trusses of the roof

Semantic analysis



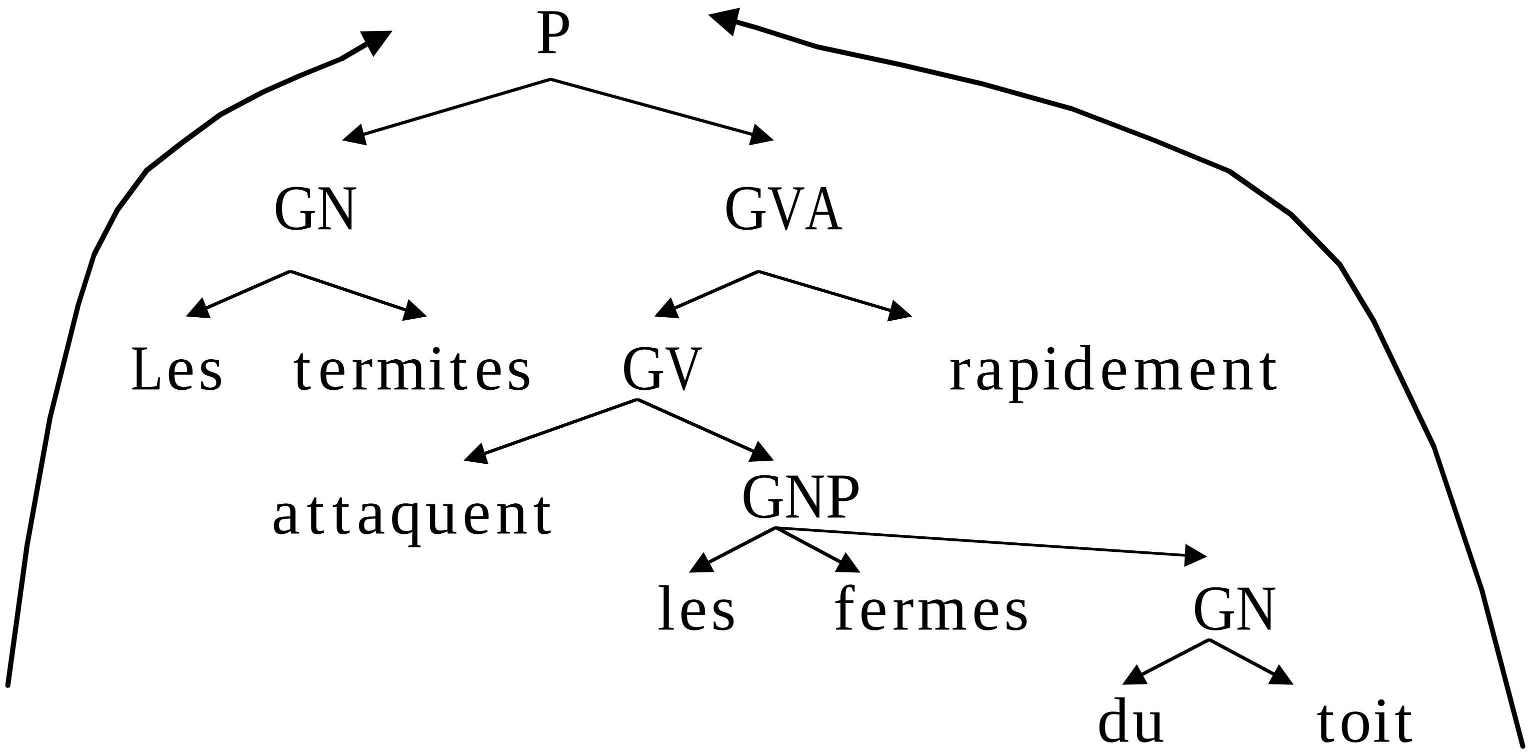
Semantic analysis

- Initialization - attach vectors to nodes



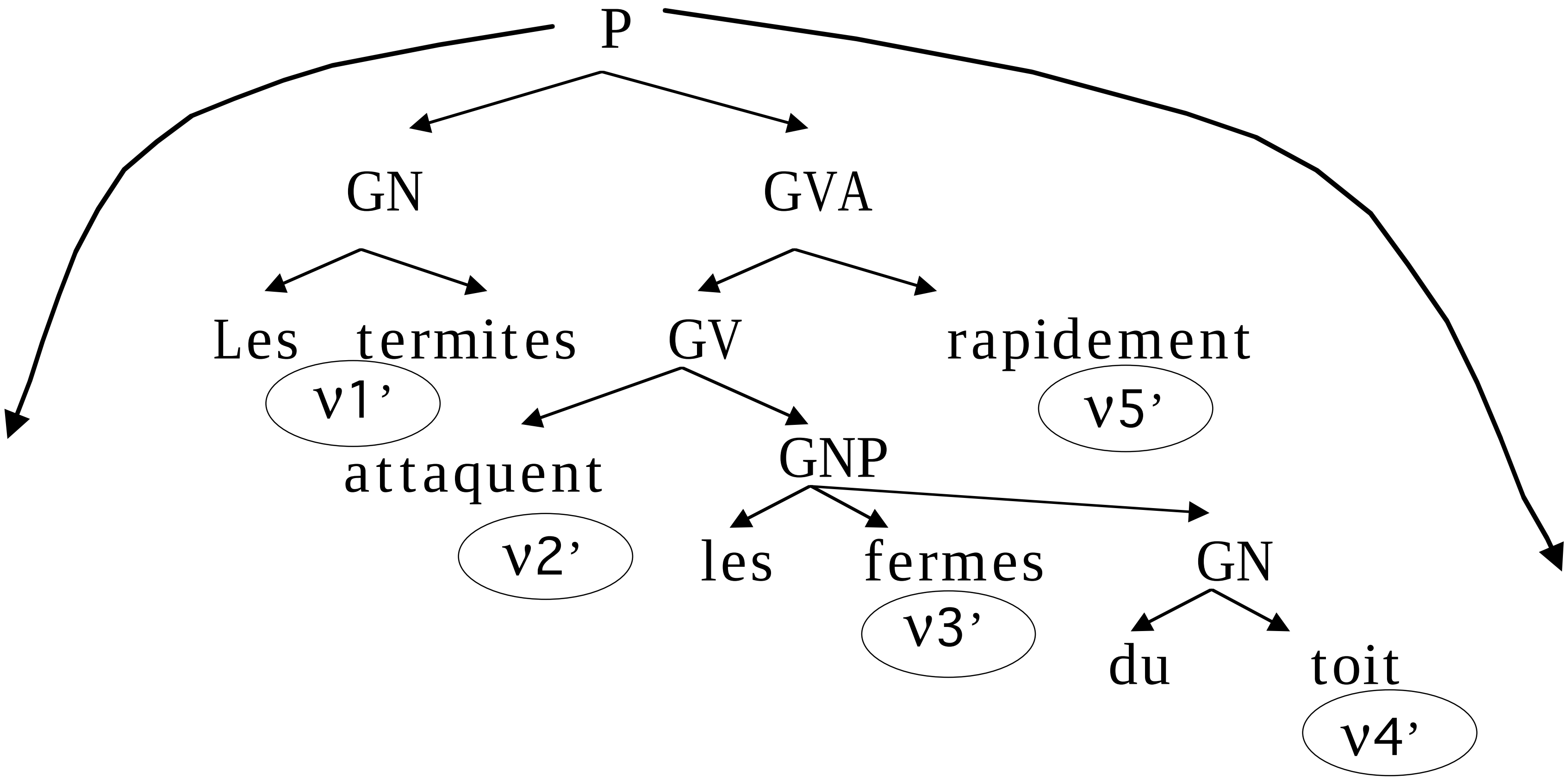
Semantic analysis

- Propagation (up)



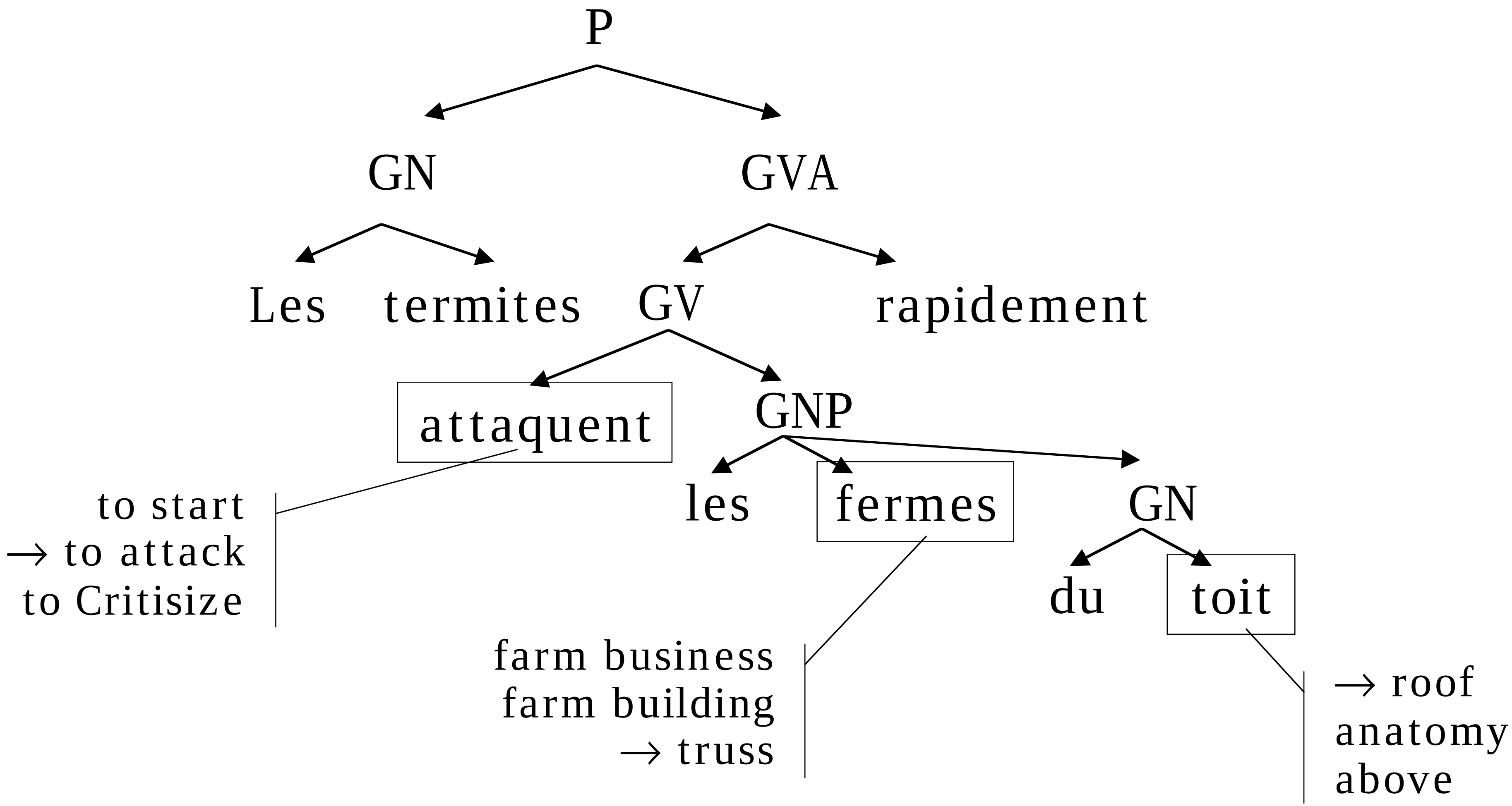
Semantic analysis

- Back propagation (down)
- $\nu(N_{i\,j}) = (\nu(N_{i\,j}) \otimes \nu(N_i)) + \nu(N_{i\,j})$



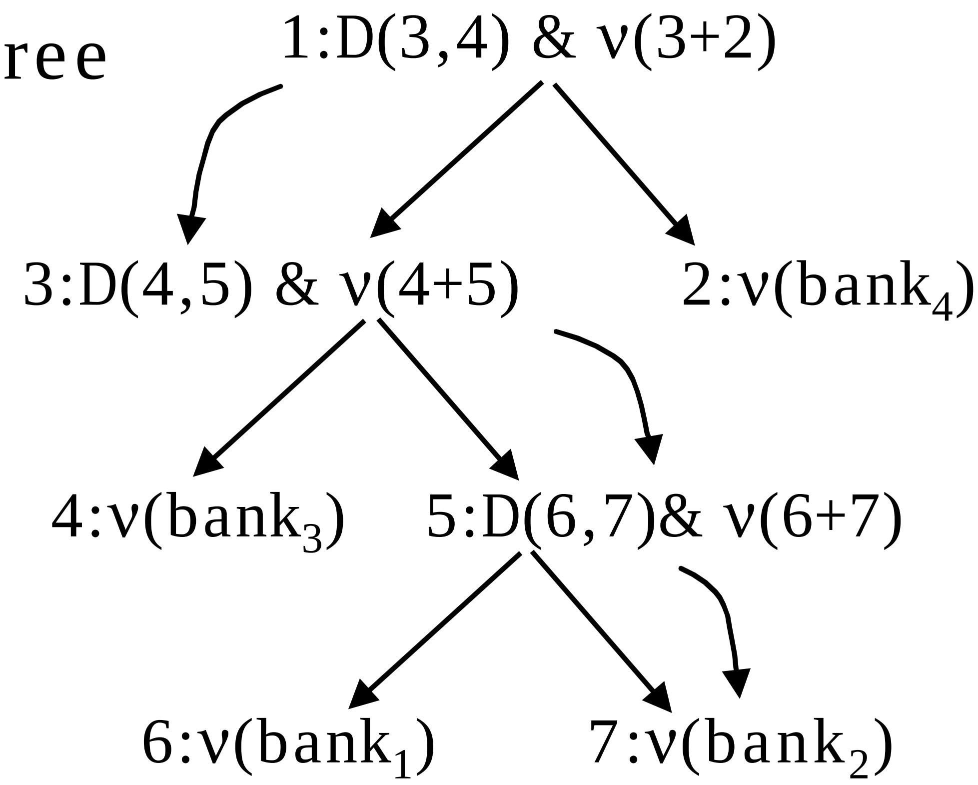
Semantic analysis

- Sense selection or sorting



Sense selection

- Recursive descent
  - on t(w) as decision tree
  - $D_A(\nu', \nu_i)$



Stop on a leaf

Stop on an internal node

Vector syntactic schemas

- S: NP(ART,N)
  - → $\nu(NP) = V(N)$

- S: NP1(NP2,N)
  - → $\nu(NP1) = \alpha\, \nu(NP1) + \nu(N)$ $0<\alpha<1$

$\nu(\text{sail boat}) = \nu(\text{sail}) + 1/2\ \nu(\text{boat})$

$\nu(\text{boat sail}) = 1/2\ \nu(\text{boat}) + \nu(\text{sail})$

Vector syntactic schemas

- Not necessary linear
- S: GA(GADV(ADV),ADJ)
  - → $\nu(GA) = \nu(ADJ)^{p(ADV)}$

  - p(very) = 2
  - p(mildly) = 1/2

ν(very happy) = $\nu(\text{happy})^2$

ν(mildly happy) = $\nu(\text{happy})^{1/2}$

Iteration & convergence

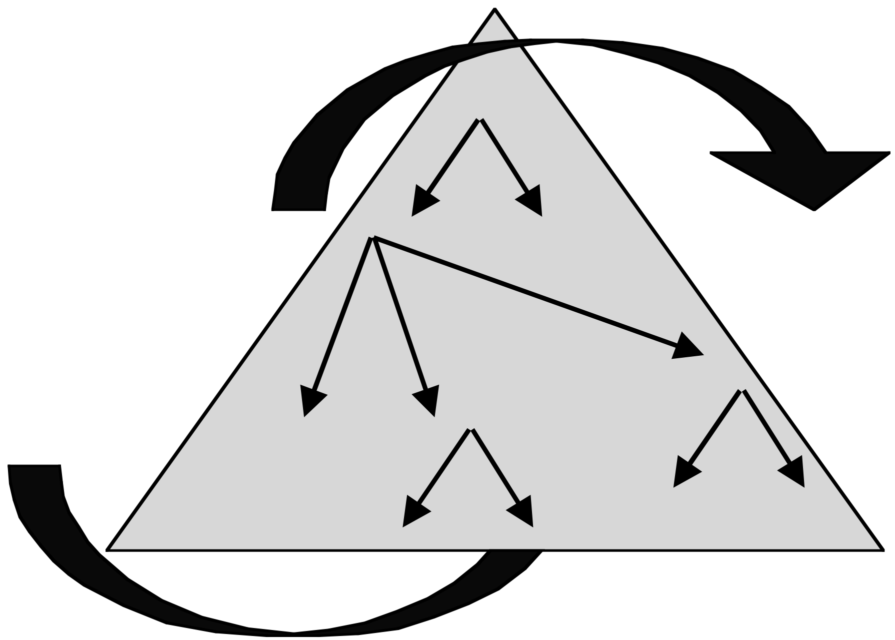
- Iteration with convergence

Local

$D(\nu_i, \nu_{i+1}) \leq \varepsilon$ *for top* ν

Global

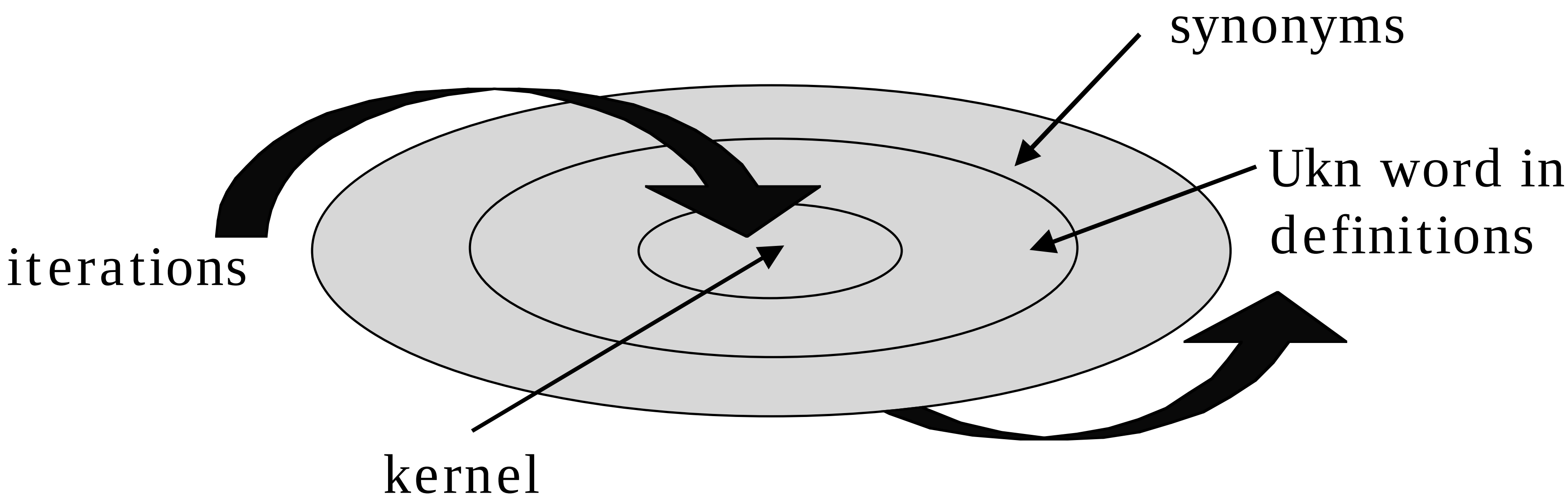
$D(\nu_i, \nu_{i+1}) \leq \varepsilon$ *for all* ν



Good results but costly

Lexicon construction

- Manual kernel
- Automatic definition analysis
- Global infinite loop = learning
- Manual adjustments



Application

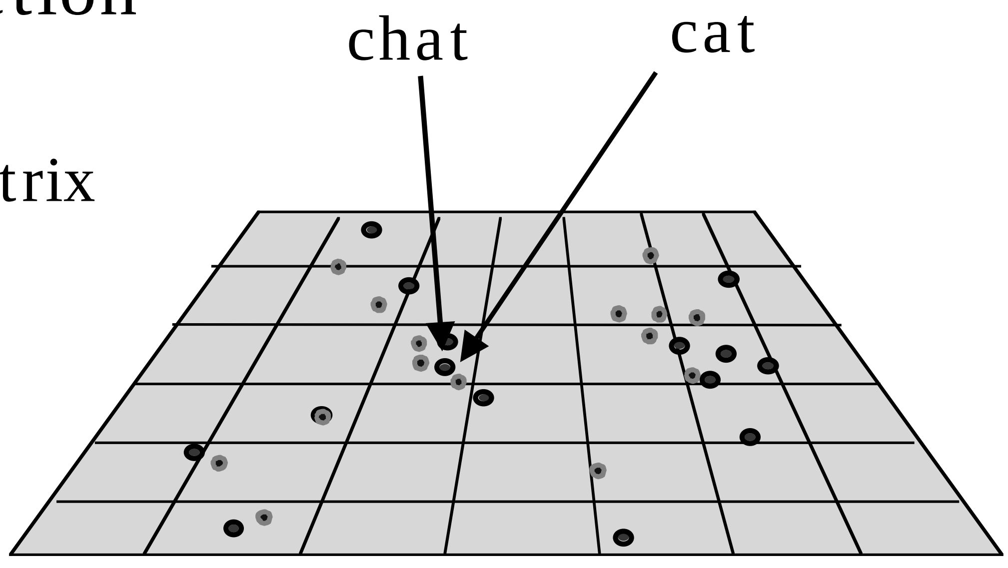
machine translation

- Lexical transfer
  - $\nu_{\text{source}} \rightarrow \nu_{\text{target}}$
  - *Knn search* that minimizes $D_A(\nu_{\text{source}}, \nu_{\text{target}})$

  - Submeaning selection

    Direct

    Transformation matrix



Application

Information Retrieval on Texts

- Textual document indexation
  - Language dependant
- Retrieval
  - Language independent - Multilingual
- Domain representation

    horse $\leftrightarrow$ equitation

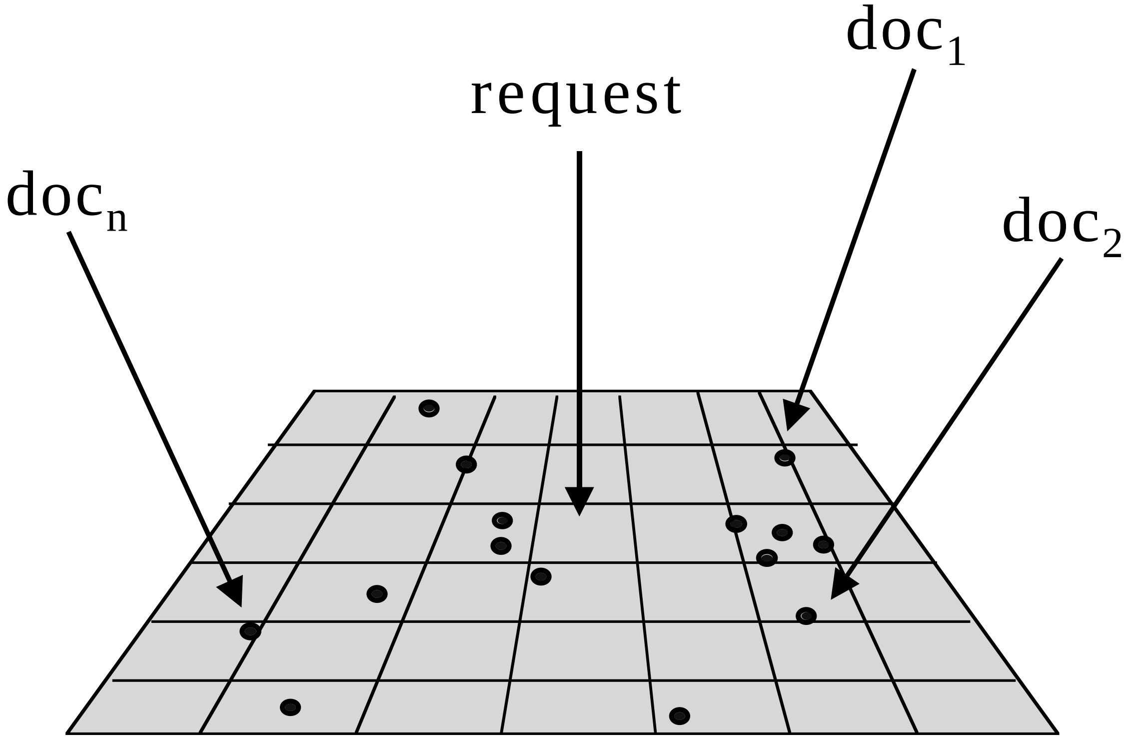
- Granularity

    Document, paragraphs, etc.

Application

Information Retrieval on Texts

- Index = Lexicon = $(d_i, \nu_i)^*$



Knn search that minimizes $D_A(\nu(r), \nu(d_i))$

Search engine

Distances adjustments

- Min $D_A(\nu(r), \nu(d_i))$ may pose problems

- Especially with small documents
  - Correlation between CV & conceptual richness
  - Pathological cases

    « plane » and « plane plane plane plane … »

    « inundation » ↔ « blood » D = 0.85 (liquid)

Search engine

Distances adjustments

- Correction with relative intensity
  - Request vs retrieved doc (ν_r and ν_d)

$$D = \sqrt{(D_A(\nu_r , \nu_d) * D_I(\nu_r , \nu_d))}$$

  - $0 \leq I(\nu_r , \nu_d) \leq 1 \quad \rightarrow \quad 0 \leq D_I(\nu_r , \nu_d) \leq \pi / 2$

Conclusion

- Approach
  - statistical (but not probabilistic)
  - thema (and rhema ?)
- Combination of
  - Symbolic methods (IA)
  - Transformational systems
- Similarity
  - Neural nets
  - With large Dim (> 50000 ?)

Conclusion

- Self evaluation
  - Vector quality
  - Tests against corpora
- Unknown words
  - Proper nouns of person, products, etc.
    - Lionel Jospin, Danone, Air France
  - Automatic learning
- Badly handled phenomena?
  - Negation & Lexical functions (Meltchuk)

End

1. extremity
2. death
3. aim

...