From Natural Language to RDF Graphs with Pregroups

Antonin Delpeuch[†], Anne Preller[*]

April 27, 2014

[†] ÉCOLE NORMALE SUPÉRIEURE

[*] Laboratoire Informatique Robotique Microélectronique Montpellier

Outline

1. Pregroup grammars and compact closed categories
2. Compositional semantics
3. Intensional model in the RDF fragment

Outline

1. Pregroup grammars and compact closed categories
2. Compositional semantics
3. Intensional model in the RDF fragment

Context

Pregroup grammars (Lambek, 1993, Lambek, 1999)

the	film	that	Emily	directed
np/n	n	$(n\backslash n)/(s/np)$	np	$(np\backslash s)/np$
$np \cdot n^l$	n	$n^r \cdot n \cdot np^{ll} \cdot s^l$	np	$np^r \cdot s \cdot np^l$

Context

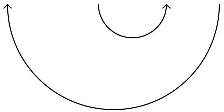
Pregroup grammars (Lambek, 1993, Lambek, 1999)

the	film	that	Emily	directed
np/n	n	$(n\backslash n)/(s/np)$	np	$(np\backslash s)/np$
$np \cdot n^l$	n	$n^r \cdot n \cdot np^{ll} \cdot s^l$	np	$np^r \cdot s \cdot np^l$

Context

Pregroup grammars (Lambek, 1993, Lambek, 1999)

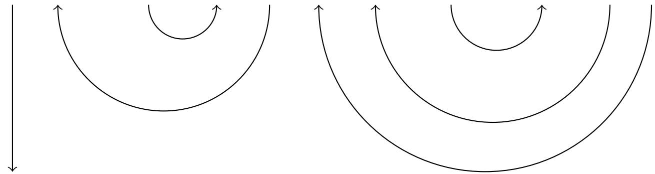
the	film	that	Emily	directed
np/n	n	$(n\backslash n)/(s/np)$	np	$(np\backslash s)/np$
$np \cdot n^l$	n	$n^r \cdot n \cdot np^{ll} \cdot s^l$	np	$np^r \cdot s \cdot np^l$



Context

Pregroup grammars (Lambek, 1993, Lambek, 1999)

the	film	that	Emily	directed
np/n	n	$(n\backslash n)/(s/np)$	np	$(np\backslash s)/np$
$np \cdot n^l$	n	$n^r \cdot n \cdot np^{ll} \cdot s^l$	np	$np^r \cdot s \cdot np^l$



Compact closed categories

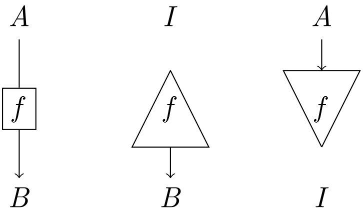
- Objects (= types)
  - are closed under $\_ \otimes \_$ (product of types), $\_^l$ and $\_^r$ (adjoints).
  - contain basic types, and I, neutral for $\otimes$.
- Arrows (= type reductions) between two objects
  - can be composed with $\circ$ (sequential composition) and $\otimes$ (parallel composition) ;
  - contain $1_A : A \to A$ (identity of A) and

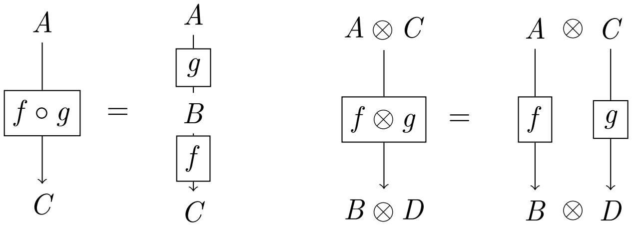
$$\epsilon^l : A^l \otimes A \to I \qquad \epsilon^r : A \otimes A^r \to I$$

$$\eta^l : I \to A \otimes A^l \qquad \eta^r : I \to A^r \otimes A$$

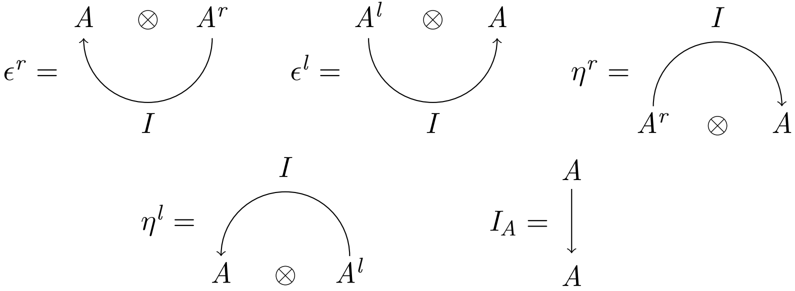
and such that some equations hold.

Representation

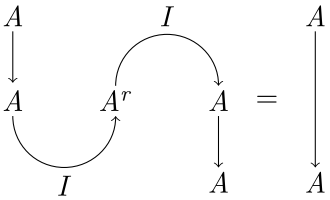


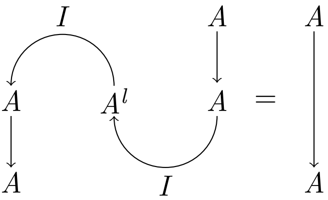


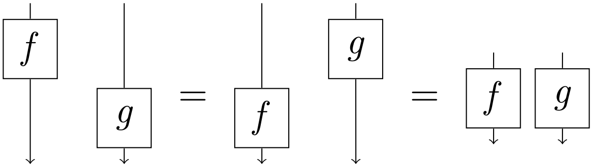
ϵ and η



Some equalities



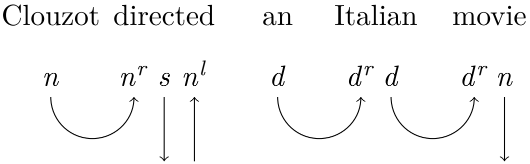




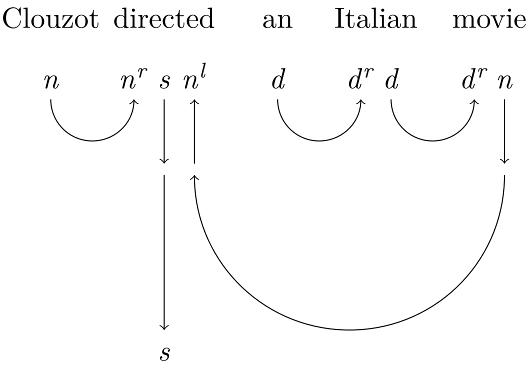
Pregroup reductions as arrows

Clouzot	directed	an	Italian	movie
n	$n^r\ s\ n^l$	d	$d^r\ d$	$d^r\ n$

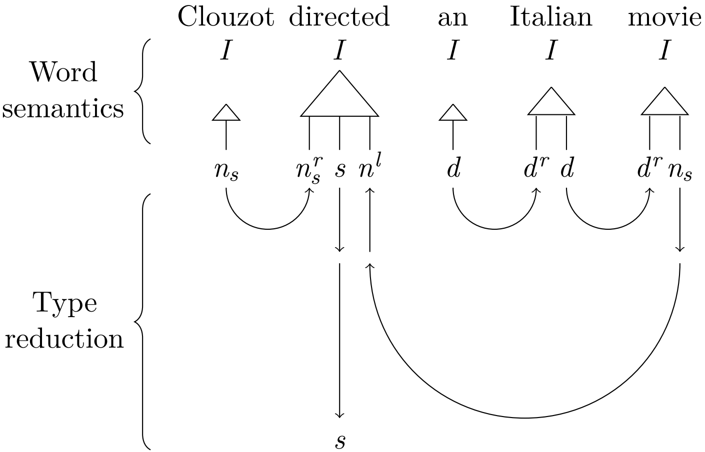
Pregroup reductions as arrows



Pregroup reductions as arrows



Compositional semantics



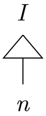
Motto: Type reduction ∘ Word meanings = Sentence meaning

Outline

1. Pregroup grammars and compact closed categories
2. Compositional semantics
3. Intensional model in the RDF fragment

Distributional semantics

Category: Finite dimensional Hilbert spaces and linear maps



$$= \begin{pmatrix} 0.73 \\ -2.3 \\ 0.1 \\ 1.4 \end{pmatrix}$$

Coecke, Sadrzadeh, and Clark (2010)
Grefenstette and Sadrzadeh (2011)

Montagovian semantics

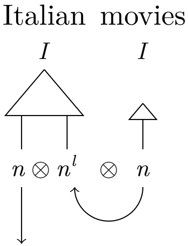
Category: Finite dimensional modules over $\{0, 1\}$ and linear maps

$$= \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

Preller and Sadrzadeh (2011)

Details

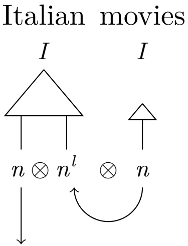
n: vector space with an infinite basis B. Each object corresponds to a $b \in B$.



Details

n: vector space with an infinite basis B. Each object corresponds to a $b \in B$.

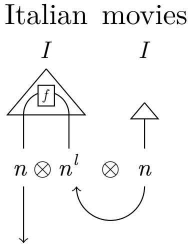
movies : $I \to n$
movies $= \sum_x \delta_{x \text{ is a movie}} x$



Details

n: vector space with an infinite basis B. Each object corresponds to a $b \in B$.

$$\text{movies} : I \to n$$
$$\text{movies} = \sum_x \delta_{x \text{ is a movie}} x$$



$$f : n \to n$$
$$f(e) = \delta_{e \text{ is Italian}} e$$

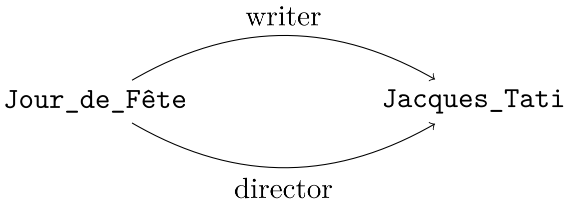
$$f(\text{movies}) = \sum_x \delta_{x \text{ is a movie}} \delta_{x \text{ is Italian}} x$$

Outline

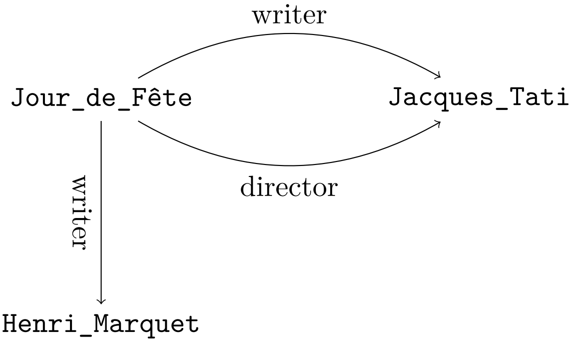
Resource Description Framework (RDF)

$$\texttt{Jour\_de\_Fête} \xrightarrow[\text{director}]{} \texttt{Jacques\_Tati}$$

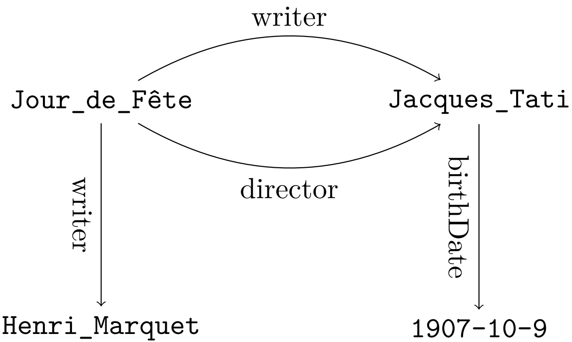
Resource Description Framework (RDF)



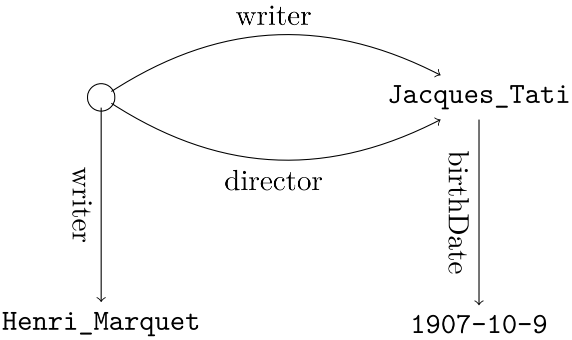
Resource Description Framework (RDF)



Resource Description Framework (RDF)



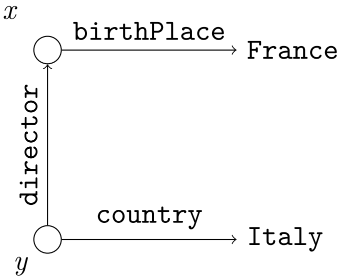
Resource Description Framework (RDF)



Representation

A French director made a movie in Italy.

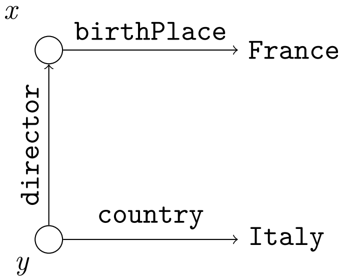
$$\exists x \exists y,\ \text{French}(x) \wedge \text{directed}(x, y) \wedge \text{in}(x, \text{Italy})$$



Representation

A French director made a movie in Italy.

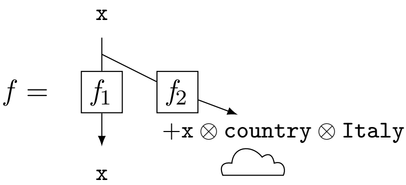
$$\exists x \exists y,\ \text{French}(x) \wedge \text{directed}(x, y) \wedge \text{in}(x, \text{Italy})$$

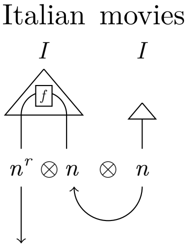


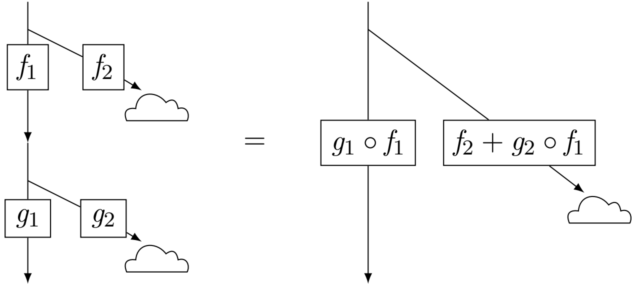
$$\texttt{x} \otimes \texttt{birthPlace} \otimes \texttt{France}$$
$$+\ \texttt{y} \otimes \texttt{director} \otimes \texttt{x}$$
$$+\ \texttt{y} \otimes \texttt{country} \otimes \texttt{Italy}$$

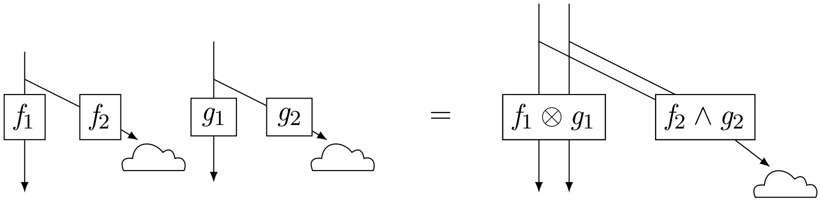
Side effects

Idea: enable arrows to add triples to a global graph in a *side effect* fashion.









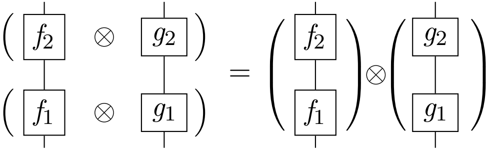
where $f_2 \wedge g_2 : e_i \otimes e_j \mapsto f_2(e_i) + g_2(e_j)$

Exchange law

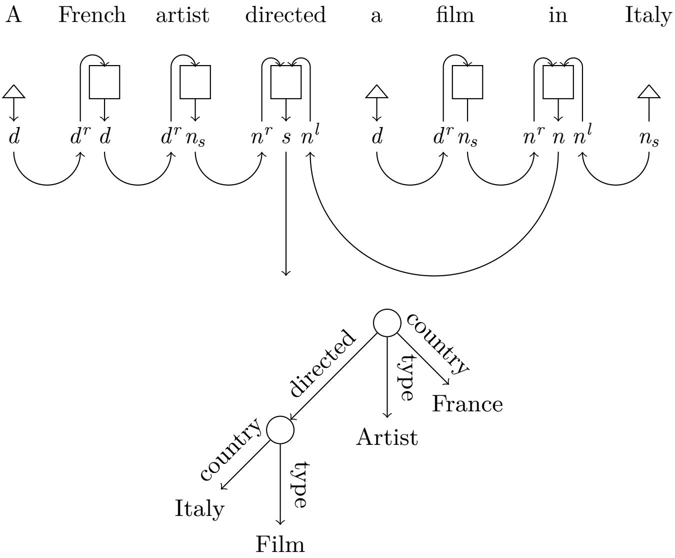
Lemma

This category is monoidal.

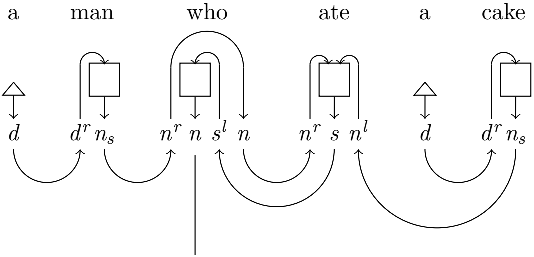
In other words, we have the following equality:

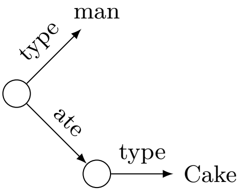


Example 1: adjectives and verbs

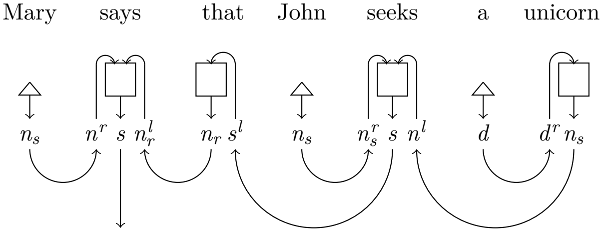


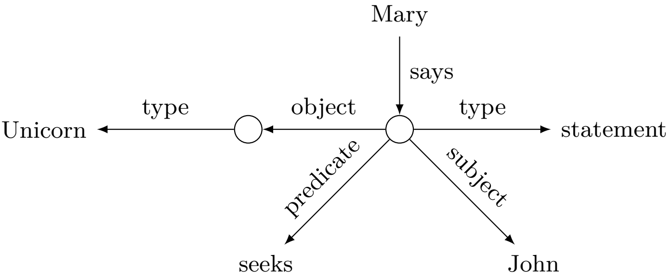
Example 2: who





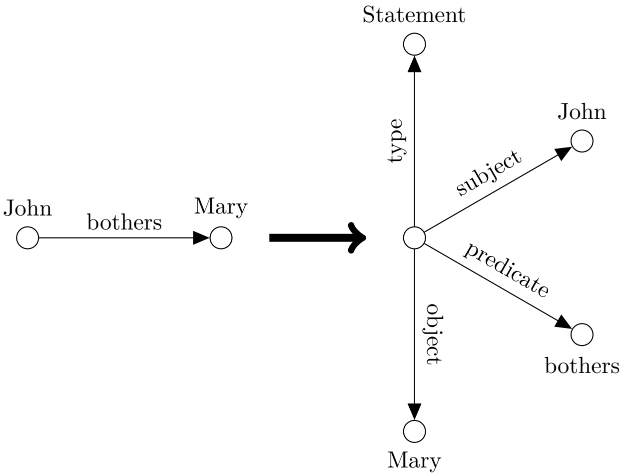
Example 3: reification





RDF reification

Reify: transform a triple into a node.



Thanks

Thanks to the TEXTE group at LIRMM, Alain Lecomte, David Naccache, Antoine Amarilli and Hugo Vanneuville.

And thank you for your attention!

Construction

The category $\mathcal{C}[S]$ has:

- objects A where A is an object in $\mathcal{C}$
- morphisms $(f, g) : A \to B$ where $f \in \mathcal{C}(A, B)$ and $g \in \mathcal{C}(A, S)$.
- a law $\otimes$ on objects as in $\mathcal{C}$
- a law $\otimes$ on arrows defined by $(f, g) \otimes (h, k) = (f \otimes h, g \wedge k)$, where $g \wedge k : u \otimes v \mapsto g(u) + k(v)$
- a law $\circ$ defined by $(f, g) \circ (h, k) = (f \circ h, g \circ h + k)$.
- arrows $1_A = (1_A, 0)$, $\epsilon^l_A = (\epsilon^l_A, 0)$, and similarly for ϵ^r, η^l and η^r.

Yanked meaning