Syntactic Control for Compositional Vector Space Models

Michael Moortgat[1] Gijs Wijnholds[2]

Utrecht Institute of Linguistics
`M.J.Moortgat@uu.nl`

ILLC Amsterdam
`gijswijnholds@gmail.com`

TACL 2015
June 21, 2015

Lambek's program

- grammar: universal type logic + language-specific lexicon
- type logic: 'deductive system as category'
- base system: biclosed tensor category (non-ass, non-comm!)
- add-ons: postulates attributing extra properties to $\otimes$:
  - unit
  - associativity
  - commutativity

The Lambek Hierarchy

$$\frac{}{1_A : A \to A}\ Id \qquad \frac{f : A \to B \quad g : B \to C}{g \circ f : A \to C}\ T$$

- **NL**:

$$\frac{f : A \otimes B \to C}{\rhd f : A \to C/B}\ Res \qquad \frac{f : A \otimes B \to C}{\lhd f : B \to A\backslash C}\ Res$$

The Lambek Hierarchy

$$\frac{}{1_A : A \to A}\ Id \qquad \frac{f : A \to B \quad g : B \to C}{g \circ f : A \to C}\ T$$

- **NL**:

$$\frac{f : A \otimes B \to C}{\rhd f : A \to C/B}\ Res \qquad \frac{f : A \otimes B \to C}{\lhd f : B \to A\backslash C}\ Res$$

- **L**: **NL** $+$

$$a_{A,B,C} : (A \otimes B) \otimes C \longleftrightarrow A \otimes (B \otimes C) : a^{-1}_{A,B,C}$$

The Lambek Hierarchy

$$\frac{}{1_A : A \to A}\ Id \qquad \frac{f : A \to B \quad g : B \to C}{g \circ f : A \to C}\ T$$

- **NL**:

$$\frac{f : A \otimes B \to C}{\rhd f : A \to C/B}\ Res \qquad \frac{f : A \otimes B \to C}{\lhd f : B \to A\backslash C}\ Res$$

- **L**: **NL** $+$

$$a_{A,B,C} : (A \otimes B) \otimes C \longleftrightarrow A \otimes (B \otimes C) : a^{-1}_{A,B,C}$$

- **LP**: **L** $+$

$$c_{A,B} : A \otimes B \to B \otimes A$$

The Lambek Hierarchy

$$\frac{}{1_A : A \to A}\ Id \qquad \frac{f : A \to B \quad g : B \to C}{g \circ f : A \to C}\ T$$

- **NL**:
$$\frac{f : A \otimes B \to C}{\rhd f : A \to C/B}\ Res \qquad \frac{f : A \otimes B \to C}{\lhd f : B \to A\backslash C}\ Res$$
- **L**: **NL** $+$
$$a_{A,B,C} : (A \otimes B) \otimes C \longleftrightarrow A \otimes (B \otimes C) : a^{-1}_{A,B,C}$$
- **LP**: **L** $+$
$$c_{A,B} : A \otimes B \to B \otimes A$$
- **U(N)L(P)**: add unit I
$$l_A : I \otimes A \longleftrightarrow A : l_A^{-1} \qquad r_A : A \otimes I \longleftrightarrow A : r_A^{-1}$$

The Lambek Hierarchy: logics, categories

$$\begin{tikzcd}
 & \mathbf{(U)LP} & & & \mathbf{SymMon} & \\
\mathbf{(U)NLP} \arrow[ur, "+\ \mathsf{a}"] & & \mathbf{(U)L} \arrow[ul, "+\ \mathsf{c}"'] \arrow[r, red, "\equiv"] & \mathbf{SymTen} \arrow[ur, "+\ \alpha"] & & \mathbf{AssTen} \arrow[ul, "+\ \sigma"'] \\
 & \mathbf{(U)NL} \arrow[ul, "+\ \mathsf{c}"] \arrow[ur, "+\ \mathsf{a}"'] & & & \mathbf{Ten} \arrow[ul, "+\ \sigma"] \arrow[ur, "+\ \alpha"'] &
\end{tikzcd}$$

- **Ten** : bi-closed tensor category
- **Mon** : bi-closed monoidal category
- **Sym**, **Ass** respectively symmetry and associativity.

Lambek Grammars

- Obtained by adding a lexicon to the logic:

$$\begin{array}{lcl} \text{john} & : & np \\ \text{mary} & : & np \\ \text{loves} & : & (np\backslash s)/np \end{array}$$

- a sequence of words $w_1 \cdots w_n$ is a well-formed expression of type B just in case the lexicon assigns types A_i to w_i and there is a proof $A^{\otimes} \rightarrow B$, where $A^{\otimes}$ is some product of the A_i

$$\dfrac{\dfrac{\dfrac{\dfrac{}{1_{np} : np \rightarrow np}\,Id \qquad \dfrac{\dfrac{}{1_{np} : np \rightarrow np}\,Id \quad \dfrac{}{1_s : s \rightarrow s}\,Id}{1_{np}\backslash 1_s : np\backslash s \rightarrow np\backslash s}\,M\backslash}{(1_{np}\backslash 1_s)/1_{np} : (np\backslash s)/np \rightarrow (np\backslash s)/np}\,M/}{\triangleright^{-1}((1_{np}\backslash 1_s)/1_{np}) : (np\backslash s)/np \otimes np \rightarrow np\backslash s}\,r}{\triangleleft^{-1}\,\triangleright^{-1}\,((1_{np}\backslash 1_s)/1_{np}) : np \otimes (((np\backslash s)/np) \otimes np) \rightarrow s}\,r$$

Compositional interpretation

- Montague: homomorphism relating source/target algebras
- Lambek: functorial transition from source to target category

$$\begin{tikzcd} \boxed{\begin{matrix}\text{Source}\\ \mathcal{C}\end{matrix}} \arrow[r, "I"] & \boxed{\begin{matrix}\text{Target}\\ \mathcal{D}\end{matrix}} \end{tikzcd}$$

Compositional interpretation

- Montague: homomorphism relating source/target algebras
- Lambek: functorial transition from source to target category

$$\begin{tikzcd}\boxed{\begin{array}{c}\text{Source}\\ \mathcal{C}\end{array}} \arrow[r, "I"] & \boxed{\begin{array}{c}\text{Target}\\ \mathcal{D}\end{array}}\end{tikzcd}$$

Target in standard Montague Grammar: set-theoretic semantics, built from universe E, truth values $\{0, 1\}$.

Compositional interpretation

- Montague: homomorphism relating source/target algebras
- Lambek: functorial transition from source to target category

$$\begin{tikzcd} \boxed{\begin{array}{c}\text{Source}\\ \mathcal{C}\end{array}} \arrow[r, "I"] & \boxed{\begin{array}{c}\text{Target}\\ \mathcal{D}\end{array}} \end{tikzcd}$$

Target in standard Montague Grammar: set-theoretic semantics, built from universe E, truth values $\{0, 1\}$.

- E.g. $I(man) = \{e \in E \mid e \text{ is a man}\}$

Compositional interpretation

- Montague: homomorphism relating source/target algebras
- Lambek: functorial transition from source to target category

$$\begin{tikzcd} \boxed{\begin{array}{c}\text{Source}\\ \mathcal{C}\end{array}} \arrow[r, "I"] & \boxed{\begin{array}{c}\text{Target}\\ \mathcal{D}\end{array}} \end{tikzcd}$$

Target in standard Montague Grammar: set-theoretic semantics, built from universe E, truth values $\{0, 1\}$.

- E.g. $I(man) = \{e \in E \mid e \text{ is a man}\}$
- We now have $I(\lhd^{-1} \rhd^{-1} ((1_{np} \backslash 1_s)/1_{np})) = \lambda a \otimes F \otimes b.F\ b\ a$ and $I(john\ loves\ mary) = \mathbf{j} \otimes \mathbf{love} \otimes \mathbf{m}$, so we get:

$$[\![john\ loves\ mary]\!] = \lambda a \otimes F \otimes b.F\ b\ a(\mathbf{j} \otimes \mathbf{love} \otimes \mathbf{m}) = \mathbf{love}(\mathbf{m})(\mathbf{j})$$

which will be 1 precisely when **j** and **m** are in the love relation.

Vector Space Models

- In the Montagovian world, semantics is limited to binary truth ↯ .
- In general, the target structure can be any bi-closed tensor/monoidal category.
- Adopting the “meaning as use” view, we would like to use finite-dimensional vector spaces as a target structure.
- The category **FVect** is in fact a bi-closed monoidal category.
- The meaning of words is induced from a corpus by means of co-occurrence counts.
- One can now reason about meaning *similarity*.

Cf. [Coecke et al 2010, 2013]

Illustrating Vector Space Models

We use two basic spaces N, S spanned by entities and $\overrightarrow{1}$, resp:

- We enhance the running example with vectors:

$$\begin{array}{lll} \text{john} & : \quad np & : \quad \overrightarrow{n_{13}} \in N \\ \text{mary} & : \quad np & : \quad \overrightarrow{n_7} \in N \\ \text{loves} & : \quad (np\backslash s)/np & : \quad .8(\overrightarrow{n_{13}} \otimes \overrightarrow{1} \otimes \overrightarrow{n_7}) + .2(\overrightarrow{n_7} \otimes \overrightarrow{1} \otimes \overrightarrow{n_{13}}) \\ & & \quad + \sum_{ij} w_{ij}(\overrightarrow{n_i} \otimes \overrightarrow{1} \otimes \overrightarrow{n_j}) \in N \otimes S \otimes N \end{array}$$

- Application is now computed using inner products:

$$\overrightarrow{n_{13}} \otimes .8(\overrightarrow{n_{13}} \otimes \overrightarrow{1} \otimes \overrightarrow{n_7}) + .2(\overrightarrow{n_7} \otimes \overrightarrow{1} \otimes \overrightarrow{n_{13}}) + \sum_{ij} w_{ij}(\overrightarrow{n_i} \otimes \overrightarrow{1} \otimes \overrightarrow{n_j}) \otimes \overrightarrow{n_7}$$

$$\mapsto .8\langle\overrightarrow{n_{13}}|\overrightarrow{n_{13}}\rangle\langle\overrightarrow{n_7}|\overrightarrow{n_7}\rangle\overrightarrow{1} + .2\langle\overrightarrow{n_{13}}|\overrightarrow{n_7}\rangle\langle\overrightarrow{n_{13}}|\overrightarrow{n_7}\rangle\overrightarrow{1} + \sum_{ij} w_{ij}\langle\overrightarrow{n_{13}}|\overrightarrow{n_i}\rangle\langle\overrightarrow{n_j}|\overrightarrow{n_7}\rangle\overrightarrow{1}$$

$$= .8\overrightarrow{1} + 0\overrightarrow{1} + 0\overrightarrow{1}$$

$$= .8\overrightarrow{1}$$

Illustrating Vector Space Models

We use two basic spaces N, S spanned by entities and $\overrightarrow{1}$, resp:

- We enhance the running example with vectors:

$$\begin{array}{lllll}
\text{john} & : & np & : & \overrightarrow{n_{13}} \in N \\
\text{mary} & : & np & : & \overrightarrow{n_7} \in N \\
\text{loves} & : & (np\backslash s)/np & : & .8(\overrightarrow{n_{13}} \otimes \overrightarrow{1} \otimes \overrightarrow{n_7}) + .2(\overrightarrow{n_7} \otimes \overrightarrow{1} \otimes \overrightarrow{n_{13}}) \\
 & & & & + \sum_{ij} w_{ij}(\overrightarrow{n_i} \otimes \overrightarrow{1} \otimes \overrightarrow{n_j}) \in N \otimes S \otimes N
\end{array}$$

- Application is now computed using inner products:

$$\begin{array}{l}
\overrightarrow{n_{13}} \otimes .8(\overrightarrow{n_{13}} \otimes \overrightarrow{1} \otimes \overrightarrow{n_7}) + .2(\overrightarrow{n_7} \otimes \overrightarrow{1} \otimes \overrightarrow{n_{13}}) + \sum_{ij} w_{ij}(\overrightarrow{n_i} \otimes \overrightarrow{1} \otimes \overrightarrow{n_j}) \otimes \overrightarrow{n_7} \\
\mapsto .8\langle \overrightarrow{n_{13}} | \overrightarrow{n_{13}} \rangle \langle \overrightarrow{n_7} | \overrightarrow{n_7} \rangle \overrightarrow{1} + .2\langle \overrightarrow{n_{13}} | \overrightarrow{n_7} \rangle \langle \overrightarrow{n_{13}} | \overrightarrow{n_7} \rangle \overrightarrow{1} + \sum_{ij} w_{ij} \langle \overrightarrow{n_{13}} | \overrightarrow{n_i} \rangle \langle \overrightarrow{n_j} | \overrightarrow{n_7} \rangle \overrightarrow{1} \\
= .8\overrightarrow{1} + 0\overrightarrow{1} + 0\overrightarrow{1} \\
= .8\overrightarrow{1}
\end{array}$$

- Question: Who is the lucky one?

Problems with the standard Lambek systems

Overgeneration because of global structural postulates

- global Associativity: grammaticality is preserved under arbitrary restructuring ↯
- global Commutativity: grammaticality is preserved under arbitrary reordering ↯

Undergeneration for logics respecting word order

- recognizing capacity **(N)L**: strictly context-free
- attested patterns beyond CF: w^2, $a^n b^n c^n$, . . .

Extended type logics

Three stategies for dealing with over/undergeneration issues:

- unary modalities for structural control: $\Diamond, \Box$
- arrow reversal symmetry: $\otimes, \oplus$
- primitive (pre/post) negations instead of units: ${}^{\mathbf{0}}\cdot, \cdot^{\mathbf{0}}, {}^{\mathbf{1}}\cdot, \cdot^{\mathbf{1}}$

Unary modalities for structural control

- We add a pair of unary connectives $\Diamond, \Box$ satisfying

$$\frac{\Diamond A \rightarrow B}{A \rightarrow \Box B}\ rp$$

- Logic: $\Diamond, \Box$ form a residuated pair. One easily shows

$$\Diamond\Box A \rightarrow A \rightarrow \Box\Diamond A \text{ (compositions)}$$

$$\text{from } A \rightarrow B \text{ infer } \Diamond A \rightarrow \Diamond B,\ \Box A \rightarrow \Box B \text{ (mon)}$$

- Structure: *global* rules $\rightsquigarrow$ $\Diamond$ controlled *restricted* versions, e.g.

$$\begin{array}{ll} A^{\diamond}: & (A \otimes B) \otimes \Diamond C \rightarrow A \otimes (B \otimes \Diamond C) \\ C^{\diamond}: & (A \otimes B) \otimes \Diamond C \rightarrow (A \otimes \Diamond C) \otimes B \end{array}$$

Illustration: non-peripheral extraction

$\Diamond\Box np$: ‘moveable’ np; key-and-lock: $\Diamond\Box np \to np$, once in place.

$$
\frac{\overset{\text{what}}{np/(s/\Diamond\Box np)} \qquad \cfrac{\cfrac{\cfrac{\cfrac{\overset{\text{Alice}}{np} \quad \cfrac{\cfrac{\overset{\text{found}}{(np\backslash s)/np} \quad \Diamond\Box np \to np}{\text{found} \otimes \Diamond\Box np \to np\backslash s}}{(\text{found} \otimes \Diamond\Box np) \otimes \text{there} \to np\backslash s} \quad \overset{\text{there}}{(np\backslash s)\backslash(np\backslash s)}}{\text{Alice} \otimes ((\text{found} \otimes \Diamond\Box np) \otimes \text{there}) \to s}}{\text{Alice} \otimes ((\text{found} \otimes \text{there}) \otimes \Diamond\Box np) \to s}\, C^{\Diamond}}{(\text{Alice} \otimes (\text{found} \otimes \text{there})) \otimes \Diamond\Box np \to s}\, A^{\Diamond}}{\text{Alice} \otimes (\text{found} \otimes \text{there}) \to s/\Diamond\Box np}}{\text{what} \otimes (\text{Alice} \otimes (\text{found} \otimes \text{there})) \to np}
$$

Compare: 'Alice found *something interesting* there'

A general theory of structural control

Let $\mathcal{L}' = \mathcal{L} + P$ for some structural postulate P (Ass, Comm). Kurtonina & Moortgat 1997: two types of modal translation to relate $\mathcal{L}, \mathcal{L}'$:

- $\mathcal{L}_{/,\otimes,\backslash} \vdash A \rightarrow B$ iff $\mathcal{L}'_{\Diamond,\Box,/,\otimes,\backslash} \vdash A^\flat \rightarrow B^\flat$

  inhibiting: $\cdot^\flat$ blocks applicabity of structural option P
- $\mathcal{L}'_{/,\otimes,\backslash} \vdash A \rightarrow B$ iff $\mathcal{L}_{\Diamond,\Box,/,\otimes,\backslash} + P_\diamond \vdash A^\sharp \rightarrow B^\sharp$

  licensing: $\cdot^\sharp$ provides access to a *controlled* version of P

We illustrate with **NL** vs **L**.

Controlling $\otimes$ Associativity

One schema serves for the licensing/inhibiting directions:

$$\begin{array}{rcl}
p^{\natural} & = & p \\
(A \otimes B)^{\natural} & = & \Diamond(A^{\natural} \otimes B^{\natural}) \\
(A/B)^{\natural} & = & \Box A^{\natural}/B^{\natural} \\
(B\backslash A)^{\natural} & = & B^{\natural}\backslash\Box A^{\natural}
\end{array}$$

Controlling $\otimes$ Associativity

One schema serves for the licensing/inhibiting directions:

$$\begin{array}{rcl} p^\natural & = & p \\ (A \otimes B)^\natural & = & \Diamond(A^\natural \otimes B^\natural) \\ (A/B)^\natural & = & \Box A^\natural / B^\natural \\ (B\backslash A)^\natural & = & B^\natural \backslash \Box A^\natural \end{array}$$

- expressing **NL** in **L**: $\Diamond$ blocks applicability of Ass, e.g.

$$\not\vdash ((a\backslash b) \otimes (b\backslash c))^\flat \rightarrow (a\backslash c)^\flat$$

Controlling ⊗ Associativity

One schema serves for the licensing/inhibiting directions:

$$\begin{array}{rcl} p^{\natural} & = & p \\ (A \otimes B)^{\natural} & = & \Diamond(A^{\natural} \otimes B^{\natural}) \\ (A/B)^{\natural} & = & \Box A^{\natural}/B^{\natural} \\ (B\backslash A)^{\natural} & = & B^{\natural}\backslash\Box A^{\natural} \end{array}$$

- expressing **NL** in **L**: ◊ blocks applicability of Ass, e.g.

$$\nvdash ((a\backslash b) \otimes (b\backslash c))^{\flat} \to (a\backslash c)^{\flat}$$

- expressing **L** in **NL**: ◊ provides access to controlled Ass

$$\Diamond(\Diamond(A \otimes B) \otimes C) \longleftrightarrow \Diamond(A \otimes \Diamond(B \otimes C)) \quad (Ass^{\diamond}) = (Ass)^{\sharp}$$

Arrow reversal symmetry: fusion vs fission

- cotensor $\oplus$, left/right difference $\oslash$, $\oslash$ satisfying

$$\frac{C \rightarrow B \oplus A}{B \obslash C \rightarrow A}\ dr \qquad \frac{C \rightarrow B \oplus A}{C \oslash A \rightarrow B}\ dr$$

arrow reversal: compare $A \otimes (A \backslash B) \rightarrow B$ vs $B \rightarrow (B \oslash A) \oplus A$

Arrow reversal symmetry: fusion vs fission

- cotensor $\oplus$, left/right difference $\obslash, \oslash$ satisfying

$$\frac{C \to B \oplus A}{B \obslash C \to A}\ dr \qquad\qquad \frac{C \to B \oplus A}{C \oslash A \to B}\ dr$$

arrow reversal: compare $A \otimes (A\backslash B) \to B$ vs $B \to (B \oslash A) \oplus A$

- add-ons: linear distributivities, respecting order, structure. G1, G3: mixed Ass; G2, G4: mixed Comm.

$$\frac{A \otimes B \to C \oplus D}{C \obslash A \to D/B}\ G1 \qquad\qquad \frac{A \otimes B \to C \oplus D}{C \obslash B \to A\backslash D}\ G2$$

$$\frac{A \otimes B \to C \oplus D}{B \oslash D \to A\backslash C}\ G3 \qquad\qquad \frac{A \otimes B \to C \oplus D}{A \oslash D \to C/B}\ G4$$

(the bidirectional case requires modal control)

Illustration: Japanese questions

Question word: infix. E.g. ‘What did John buy?’
'John-ga nani-o katta ka' — lit. 'John what bought Q'

Illustration: Japanese questions

Question word: infix. E.g. ‘What did John buy?’
'John-ga nani-o katta ka’ — lit. 'John what bought Q’

$$
\begin{array}{cl}
\begin{array}{cc}
\vdots & \vdots \\ \hline
n \otimes (a \otimes (a\backslash(n\backslash s))) \rightarrow s & q \rightarrow (q \oslash wh) \oplus wh
\end{array} & \\ \hline
s\backslash q \rightarrow (n \otimes (a \otimes (a\backslash(n\backslash s))))\backslash((q \oslash wh) \oplus wh) & \\ \hline
(n \otimes (a \otimes (a\backslash(n\backslash s)))) \otimes (s\backslash q) \rightarrow (q \oslash wh) \oplus wh & \\ \hline
(q \oslash wh) \obslash (n \otimes (a \otimes (a\backslash(n\backslash s)))) \rightarrow wh/(s\backslash q) & G1 \\ \hline
n \otimes (a \otimes (a\backslash(n\backslash s))) \rightarrow (q \oslash wh) \oplus (wh/(s\backslash q)) & \\ \hline
(q \oslash wh) \obslash (a \otimes (a\backslash(n\backslash s))) \rightarrow n\backslash(wh/(s\backslash q)) & G2 \\ \hline
a \otimes (a\backslash(n\backslash s)) \rightarrow (q \oslash wh) \oplus (n\backslash(wh/(s\backslash q))) & \\ \hline
(q \oslash wh) \obslash a \rightarrow (n\backslash(wh/(s\backslash q)))/(a\backslash(n\backslash s)) & G1 \\ \hline
(n \otimes (((q \oslash wh) \obslash a) \otimes (a\backslash(n\backslash s)))) \otimes (s\backslash q) \rightarrow wh &
\end{array}
$$

q: yes/no question; wh: constituent question; a: direct object

Negations as Primitives

(Co)tensor units lead to overgeneration (because of $I \rightarrow A/A$ etc)

- We add primitive (pre/post) negations ${}^{0}\cdot, \cdot^{0}, {}^{1}\cdot, \cdot^{1}$ satisfying

$$\frac{A \rightarrow B^{0}}{B \rightarrow {}^{0}A}\ gc \qquad \frac{{}^{1}A \rightarrow B}{B^{1} \rightarrow A}\ dgc$$

Negations as Primitives

(Co)tensor units lead to overgeneration (because of $I \to A/A$ etc)

- We add primitive (pre/post) negations ${}^0\cdot, \cdot^0, {}^1\cdot, \cdot^1$ satisfying

$$\frac{A \to B^0}{B \to {}^0A}\ gc \qquad\qquad \frac{{}^1A \to B}{B^1 \to A}\ dgc$$

- Logic: (dual) Galois connected pairs. One easily shows

  $A \to {}^{\mathbf{0}}(A^{\mathbf{0}}),\ A \to ({}^{\mathbf{0}}A)^{\mathbf{0}}$ and ${}^{\mathbf{1}}(A^{\mathbf{1}}) \to A,\ ({}^{\mathbf{1}}A)^{\mathbf{1}} \to A$

  antitonicity: from $A \to B$ infer ${}^{\mathbf{0}}B \to {}^{\mathbf{0}}A$, etc

Negations as Primitives

(Co)tensor units lead to overgeneration (because of $I \rightarrow A/A$ etc)

- We add primitive (pre/post) negations ${}^{0}\cdot, \cdot^{0}, {}^{1}\cdot, \cdot^{1}$ satisfying

$$\frac{A \rightarrow B^{0}}{B \rightarrow {}^{0}A}\ gc \qquad \frac{{}^{1}A \rightarrow B}{B^{1} \rightarrow A}\ dgc$$

- Logic: (dual) Galois connected pairs. One easily shows

  $A \rightarrow {}^{\mathbf{0}}(A^{\mathbf{0}}),\ A \rightarrow ({}^{\mathbf{0}}A)^{\mathbf{0}}$ and ${}^{\mathbf{1}}(A^{\mathbf{1}}) \rightarrow A,\ ({}^{\mathbf{1}}A)^{\mathbf{1}} \rightarrow A$

  antitonicity: from $A \rightarrow B$ infer ${}^{\mathbf{0}}B \rightarrow {}^{\mathbf{0}}A$, etc

- Structure: collapse to 2-negation or 1-negation system

$$A^{lr} \leftrightarrow A \leftrightarrow A^{rl} \qquad A \leftrightarrow A^{\perp\perp}$$

  via negation variants of the Grishin distributivities

Compositional interpretation for the extended logics

How can we adapt the functorial transition to vector-based models of meaning to the extended type logics?

- control modalities $\Diamond, \Box$
- cotensor, left/right difference $\oplus, \obslash, \oslash$
- primitive (pre/post) negations

Interpreting $\mathbf{NL}_{\Diamond,\Box}$

- We add a basic space C, and set

$$\begin{array}{rcl} I(\Diamond A) & = & C \otimes I(A) \\ I(\Box A) & = & C^* \otimes I(A) \end{array}$$

- Assuming that C expresses some *context*, the abstract meaning recipe of “what Alice found” becomes

$$\textbf{what}\ (\lambda c \otimes F.\ \textbf{found}\ (F\ c)\ \textbf{alice})$$

- So, we interpret “what” as a vector that supplies a context and a context interpretation by which to evaluate the probability by which alice found some object.
- Context: “Once upon a time there were two keys, and Alice. Alice found the first key. It is not hard to determine what Alice found.”

Illustration

- C spanned by vectors $\overrightarrow{obj}, \overrightarrow{pers}$, and assume:

  $\overrightarrow{c} = 0.66\overrightarrow{obj} + 0.33\overrightarrow{pers}$

  $\overrightarrow{F} = \overrightarrow{obj} \otimes \overrightarrow{key_1} + \overrightarrow{obj} \otimes \overrightarrow{key_2} + 0(\overrightarrow{obj} \otimes \overrightarrow{alice}) + ... + \overrightarrow{pers} \otimes \overrightarrow{alice}$

  $\overrightarrow{found} = \overrightarrow{alice} \otimes \overrightarrow{1} \otimes \overrightarrow{key_1} + 0(\overrightarrow{alice} \otimes \overrightarrow{1} \otimes \overrightarrow{key_2}) + 0(\overrightarrow{key_1} \otimes \overrightarrow{1} \otimes \overrightarrow{alice}) + ...$
- Interpreting the context against the context interpretation vector yields
  $$\overrightarrow{F\ c} = 0.66\overrightarrow{key_1} + 0.66\overrightarrow{key_2} + 0.33\overrightarrow{alice}$$
- We interpret "what" as the vector that evaluates its argument against $\overrightarrow{c}$ and $\overrightarrow{F}$:
  $$\begin{gathered}\textbf{what}\ (\lambda c \otimes F.\ \textbf{found}\ (F\ c)\ \textbf{alice}) \\ \mapsto \langle\overrightarrow{alice}|\overrightarrow{alice}\rangle\langle\overrightarrow{key_1}|0.66\overrightarrow{key_1} + 0.66\overrightarrow{key_2} + 0.33\overrightarrow{alice}\rangle\ \overrightarrow{1} + \\ 0\langle\overrightarrow{alice}|\overrightarrow{alice}\rangle\langle\overrightarrow{key_2}|0.66\overrightarrow{key_1} + 0.66\overrightarrow{key_2} + 0.33\overrightarrow{alice}\rangle\ \overrightarrow{1} + ... \\ = 0.66\ \overrightarrow{1}\end{gathered}$$

Interpreting Fission

- We can interpret the $(\oplus, \obslash, \oslash)$ family using the standard tensor product on vector spaces:

$$\begin{array}{lcl} I(B \oplus A) & = & I(B) \otimes I(A) \\ I(A \backslash B) & = & I(A)^* \otimes I(B) \\ I(A \obslash B) & = & I(A)^* \otimes I(B) \\ & \vdots & \end{array}$$

- Linear distributivities are interpreted using the symmetry and associativity transformations of **FVect**:

$$\begin{array}{lcl} g_1 : A \obslash (B \otimes C) \to (A \obslash B) \otimes C & \rightsquigarrow & \alpha^{-1}_{A^*,B,C} \\ g_2 : A \obslash (C \otimes B) \to C \otimes (A \obslash B) & \rightsquigarrow & \alpha_{C,A^*,B} \circ (\sigma_{A^*,C} \otimes id_B) \circ \alpha^{-1}_{A^*,C,B} \\ & \vdots & \end{array}$$

Cf. [Wijnholds 2014]

Avoiding collapse

Under the above interpretation, the source distinction between the $\otimes$ and $\oplus$ families collapses at the target level.
With a two-step interpretation one can avoid this collapse:

$$\begin{array}{c}\text{source}\\ \mathbf{LG}_{\otimes,\oplus}\end{array} \xrightarrow{I} \begin{array}{c}\text{continuation}\\ \text{semantics}\\ \mathbf{LP}_{\otimes,\cdot^{\perp}}\end{array} \xrightarrow{J} \begin{array}{c}\text{target}\\ \mathbf{FVect}\end{array}$$

- intermediate level: continuation semantics for **LG**, with linear $\otimes$ and implication w.r.t. response type $\perp$ ($A^{\perp} := A \multimap \perp$)
- source distinction $\otimes, \oplus$ is retained in the interpretation

$$I(A \otimes B) = I(A) \otimes I(B) \quad vs \quad I(A \oplus B) = I(A)^{\perp} \otimes I(B)^{\perp}$$

Cf [Moortgat 2009]

Interpreting Negations

- Given a dagger structure induced by $\cdot^\dagger$, we can interpret negations:

$$\begin{array}{rcl} I(A^0) & = & I(A)^\dagger \\ I(A^1) & = & I(A)^\dagger \\ & \vdots & \end{array}$$

- Again, in the transition to the target interpretation, the source distinctions between pre/post negations are lost
- Linguistic application: lexical ambiguity
- Either **FVect** or **FHilb** allow Selinger's **CPM** construction, which can be used to measure ambiguity.
- Viz. ‘Queen rules’ vs 'Queen rules England’.

Cf. [Selinger 2007, Piedeleu et al. 2015]

Summarizing, future work

We have discussed an extended vocabulary of logical constants:

positive formulas: $A \otimes B$, $A \oslash B$, $B \obslash A$; $\Diamond A, A^{\mathbf{1}}, {}^{\mathbf{1}}A$;

negative formulas: $A \oplus B$, $A \backslash B$, B / A; $\Box A, {}^{\mathbf{0}}A, A^{\mathbf{0}}$

Summarizing, future work

We have discussed an extended vocabulary of logical constants:

positive formulas: $A \otimes B$, $A \oslash B$, $B \obslash A$; $\Diamond A, A^{\mathbf{1}}, {}^{\mathbf{1}}A$;

negative formulas: $A \oplus B$, $A \backslash B$, B / A; $\Box A, {}^{\mathbf{0}}A, A^{\mathbf{0}}$

- the resulting logics have enhanced syntactic control, avoiding the over/undergeneration issues of Lambek's original systems

Summarizing, future work

We have discussed an extended vocabulary of logical constants:

positive formulas: $A \otimes B$, $A \oslash B$, $B \obslash A$; $\Diamond A, A^{\mathbf{1}}, {}^{\mathbf{1}}A$;

negative formulas: $A \oplus B$, $A \backslash B$, B / A; $\Box A, {}^{\mathbf{0}}A, A^{\mathbf{0}}$

- the resulting logics have enhanced syntactic control, avoiding the over/undergeneration issues of Lambek's original systems
- they still allow a functorial transition to vector-based models of meaning

Summarizing, future work

We have discussed an extended vocabulary of logical constants:

positive formulas: $A \otimes B,\ A \oslash B,\ B \obslash A;\ \Diamond A, A^{\mathbf{1}}, {}^{\mathbf{1}}A;$

negative formulas: $A \oplus B,\ A\backslash B,\ B/A;\ \Box A, {}^{\mathbf{0}}A, A^{\mathbf{0}}$

- the resulting logics have enhanced syntactic control, avoiding the over/undergeneration issues of Lambek's original systems
- they still allow a functorial transition to vector-based models of meaning
- in the mapping to the target, many source distinctions are lost

Summarizing, future work

We have discussed an extended vocabulary of logical constants:

positive formulas: $A \otimes B$, $A \oslash B$, $B \obslash A$; $\Diamond A, A^{\mathbf{1}}, {}^{\mathbf{1}}A$;

negative formulas: $A \oplus B$, $A \backslash B$, B / A; $\Box A, {}^{\mathbf{0}}A, A^{\mathbf{0}}$

- the resulting logics have enhanced syntactic control, avoiding the over/undergeneration issues of Lambek's original systems
- they still allow a functorial transition to vector-based models of meaning
- in the mapping to the target, many source distinctions are lost
- can one find empirical support for *enriched* quantitative models, retaining more of the source structure?

Summarizing, future work

We have discussed an extended vocabulary of logical constants:

positive formulas: $A \otimes B$, $A \oslash B$, $B \obslash A$; $\Diamond A, A^{\mathbf{1}}, {}^{\mathbf{1}}A$;

negative formulas: $A \oplus B$, $A \backslash B$, B/A; $\Box A, {}^{\mathbf{0}}A, A^{\mathbf{0}}$

- the resulting logics have enhanced syntactic control, avoiding the over/undergeneration issues of Lambek's original systems
- they still allow a functorial transition to vector-based models of meaning
- in the mapping to the target, many source distinctions are lost
- can one find empirical support for *enriched* quantitative models, retaining more of the source structure?

. . . future work

Questions?

A. Bastenhof.
Categorial Symmetry.
PhD thesis, Utrecht University, 2013.

R. Bernardi and M. Moortgat.
Continuation semantics for the Lambek-Grishin calculus.
Inf. Comput., 208(5):397–416, 2010.

B. Coecke, M. Sadrzadeh, and S. Clark.
Mathematical foundations for a compositional distributional model of meaning.
CoRR, abs/1003.4394, 2010.

B. Jacobs.
Semantics of weakening and contraction.
Annals of Pure and Applied Logic, 69(1):73 – 106, 1994.

L. Kallmeyer.
Parsing Beyond Context-Free Grammars.
Springer, 2010.

N. Kurtonina and M. Moortgat.
Structural control.
In P. Blackburn and M. de Rijke, editors, *Specifying Syntactic Structures*, pages 75–113. CSLI, Stanford, 1997.

J. Lambek.
On the calculus of syntactic types.
In R. Jacobson, editor, *Structure of language and its mathematical aspects*, pages 166–178. American Mathematical Society, 1961.

M. Moortgat.
Symmetric categorial grammar.
J. Philosophical Logic, 38(6):681–710, 2009.

R. Piedeleu, D. Kartsaklis, B. Coecke, and M. Sadrzadeh.
Open system categorical quantum semantics in natural language processing.
CoRR, abs/1502.00831, 2015.

G. Wijnholds.
Categorical foundations for extended compositional distributional models of meaning.
MSc Thesis, ILLC, University of Amsterdam, 2014.