Classical Copying versus Quantum Entanglement in Natural Language: the Case of VP-ellipsis

Gijs Jasper Wijnholds¹  Mehrnoosh Sadrzadeh¹

Queen Mary University of London, United Kingdom
g.j.wijnholds@qmul.ac.uk

SYCO 2
December 17, 2018
DISTRIBUTIONAL SEMANTICS: MEANING IN CONTEXT

...an efficient method for learning high quality distributed vector ...
COMPOSING WORD EMBEDDINGS: A CHALLENGE

- **Coordination**: dancing and running = ??
- **Quantification**: Every student likes some teacher = ??
- **Anaphora**: shaves himself = ??
- **Ellipsis**: Matt went to Croatia and Max did too = ??
VERB PHRASE ELLIPSIS

- Ellipsis is a natural language phenomenon in which part of a phrase is missing and has to be recovered from context.
- In verb phrase ellipsis, the missing part is... a verb phrase.
- There is often a marker that indicates the type of the missing part.
VERB PHRASE ELLIPSIS

- Ellipsis is a natural language phenomenon in which part of a phrase is missing and has to be recovered from context.
- In verb phrase ellipsis, the missing part is... a verb phrase.
- There is often a marker that indicates the type of the missing part.

Bob drinks a beer and Alice does too

ant VP marker
ELLIPSIS NEEDS COPYING AND MOVEMENT

Bob drinks a beer and Alice \[
\text{\_ \_ \_}
\]
does too

ANT VP

MARKER
ELLIPSIS NEEDS COPYING AND MOVEMENT

drinks a beer  drinks a beer

Bob drinks a beer  and Alice marker does too

ANT VP
ELLIPSIS NEEDS COPYING AND MOVEMENT

Bob drinks a beer and Alice does too.

ANT VP

MARKER
THE CHALLENGE: COMPOSE WORD VECTORS TO GET A MEANING REPRESENTATION FOR VP ELLIPSIS
THE BIG PICTURE

Quantum Entanglement

SOURCE $L_{\diamond, F}$ \quad \xrightarrow{\text{Functor}} \quad \text{TARGET} \quad FVec

Classical

SOURCE $L_{\diamond, F}$ \quad \xrightarrow{H_{\text{der}}} \quad \text{INTER} \quad \lambda_{\text{NL}} \quad \xrightarrow{H_{\text{lex}}} \quad \text{TARGET} \quad \lambda_{FVec_{\text{Frob}}}$
QUANTUM ENTANGLEMENT
The core of the Lambek calculus: application, co-application

\[ B \otimes B \setminus A \rightarrow A \quad A \rightarrow B \setminus (B \otimes A) \]
\[ A / B \otimes B \rightarrow A \quad A \rightarrow (A \otimes B) / B \]
\[ (A \otimes B) \otimes C \leftrightarrow A \otimes (B \otimes C) \]

Interpretation: words have types, and type-respecting embeddings

<table>
<thead>
<tr>
<th>Word</th>
<th>Type</th>
<th>embedding</th>
</tr>
</thead>
<tbody>
<tr>
<td>john</td>
<td>np</td>
<td>( \overrightarrow{john} \in N )</td>
</tr>
<tr>
<td>sleeps</td>
<td>np(\setminus s )</td>
<td>( \overrightarrow{sleep} \in N \otimes S ) ((\leftarrow) matrix)</td>
</tr>
<tr>
<td>likes</td>
<td>(np(\setminus s))/np</td>
<td>( \overrightarrow{like} \in N \otimes S \otimes N ) ((\leftarrow) cube)</td>
</tr>
<tr>
<td>beer</td>
<td>np</td>
<td>( \overrightarrow{beer} \in N )</td>
</tr>
</tbody>
</table>

(Coecke et al., 2013)
IN PICTURES

\[
A \otimes A \backslash B \to B \quad B \to A \backslash (A \otimes B)
\]

\[
B/A \otimes A \to B \quad B \to (B \otimes A)/A
\]

(Coecke et al., 2013)
IN PICTURES

\[
A \otimes A \backslash B \rightarrow B \quad B \rightarrow A \backslash (A \otimes B)
\]

\[
B / A \otimes A \rightarrow B \quad B \rightarrow (B \otimes A) / A
\]

LINEAR!!

(Coecke et al., 2013)
LAMBEK WITH CONTROL OPERATORS: $L^{\Diamond,F}$

The core of the Lambek calculus: application, co-application

$$B \otimes B \not\rightarrow A \rightarrow A \quad A \rightarrow B \not\rightarrow (B \otimes A)$$

$$A/B \otimes B \rightarrow A \quad A \rightarrow (A \otimes B)/B$$

$$(A \otimes B) \otimes C \leftrightarrow A \otimes (B \otimes C)$$

Modalities: application, co-application

$$\Diamond \Box A \rightarrow A \quad A \rightarrow \Box \Diamond A$$

- Linear logic: controlled duplication/deletion of resources via
  $! = \Diamond \Box$. Here: controlled copying, reordering

Controlled contraction, commutativity

$$A \rightarrow \Diamond A \otimes A \quad (\Diamond A \otimes B) \otimes C \rightarrow B \otimes (\Diamond A \otimes C)$$

$$\Diamond A \otimes (\Diamond B \otimes C) \rightarrow \Diamond B \otimes (\Diamond A \otimes C)$$
ILLUSTRATION

Bob /a.sc/n.sc/t.sc VP drinks a beer and Alice /m.sc/a.sc/r.sc/k.sc/e.sc/r.sc does too

\(\Diamond (np \backslash s)\)

\(np \backslash s\)

\(np\)

\(np \backslash s\)

\((s \backslash s)/s\) \(np\)

\(\Diamond (np \backslash s) \backslash (np \backslash s)\)

Bob drinks a beer and Alice does too

ANT VP MARKER
IN PICTURES

\[ A \otimes A \setminus B \rightarrow B \]

\[ B \rightarrow A \setminus (A \otimes B) \]

\[ (\Diamond A \otimes B) \otimes C \rightarrow B \otimes (\Diamond A \otimes C) \]
QUANTUM ENTANGLEMENT AND ELLIPSIS

Bob drinks a beer

and

Alice does too
QUANTUM ENTANGLEMENT AND ELLIPSIS

Bob drinks a beer and Alice also does too.
QUANTUM ENTANGLEMENT AND ELLIPSIS

\[ (\overrightarrow{Bob} \odot \overrightarrow{Alice})^T \times \overrightarrow{drinks a beer} \]
A MORE COMPLICATED CASE: SLOPPY READING

Bob loves his beer and Alice does too \( \rightsquigarrow \)
Bob loves Bob’s beer and Alice loves Bob’s beer

\[
\begin{align*}
np \quad (np\,s)/np \quad \diamond np\,(np/n) \quad n \quad (s\,s)/s \quad np \quad \diamond (np\,s)/(np\,s)
\end{align*}
\]

Bob \quad loves \quad \_ \quad his \quad beer \quad and \quad Alice \quad \_ \quad does \quad too

ANT VP \quad MARKER
A MORE COMPLICATED CASE: SLOPPY READING

Bob loves his beer and Alice does too \(\sim\rightarrow\)
Bob loves Bob’s beer and Alice loves Bob’s beer

\[
\begin{align*}
\text{Bob} & \quad \text{loves} & \quad \text{his} & \quad \text{beer} & \quad \text{and} & \quad \text{Alice} & \quad \text{does} & \quad \text{too} \\
\text{ANT VP} & & & & & & & \\
\text{MARKER} & & & & & & &
\end{align*}
\]
A MORE COMPLICATED CASE: STRICT READING

Bob loves his beer and Alice does too $\rightsquigarrow$
Bob loves Bob’s beer and Alice loves Alice’s beer

$np \ (np/s)/np \ \Diamond np/(np/n) \ n \ (s/s)/s \ np \ \Diamond (np/s)(np/s)$

Bob loves _ his beer and Alice _ does too

ANT VP MARKER
A MORE COMPLICATED CASE: STRICT READING

Bob loves his beer and Alice does too \( \sim \)
Bob loves Bob’s beer and Alice loves Alice’s beer

\[
\begin{array}{c}
\text{Bob loves his beer and Alice does too } \sim \\
\text{Bob loves Bob’s beer and Alice loves Alice’s beer}
\end{array}
\]
A More Complicated Case: Sloppy Reading

Bob loves his beer and Alice does too
A More Complicated Case: Sloppy Reading

Bob loves his beer and Alice does too.
A More Complicated Case: Sloppy Reading

Bob loves Bob’s beer and Alice loves Bob’s beer

\[ \Delta(Bob \odot Alice \odot Beer)_{ik} \text{loves}_{ijk} \]
A More Complicated Case: Strict Reading

Bob loves Bob’s beer and Alice loves Alice’s beer
A More Complicated Case: Strict Reading

Bob loves Bob’s beer and Alice loves Alice’s beer

\[ \Delta(\text{Bob} \diamond \text{Alice} \diamond \text{Beer})_{ik} \text{loves}_{ijk} \]
WHAT NOW?
WHAT NOW?

Classical Semantics
General Interpretation

The syntax-semantics homomorphism interprets types and proofs of $\mathbf{L}^{\diamond, F}$ as objects types and maps terms in a compact closed category non-linear lambda calculus:

Type Level

$$[A \otimes B] = [A] \times [B] \quad [A/B] = [A] \to [B] \quad [A\setminus B] = [A] \to [B] \quad [\diamond A] = [\Box A] = [A]$$

Application, co-application

$$B \times (B \to A) \xrightarrow{\lambda x.M.Mx} \lambda x.\lambda y.\langle y, x \rangle$$

$$B \to (B \times A)$$

$$(B \to A) \times B \xrightarrow{\lambda Mx.Mx} \lambda x.\lambda y.\langle x, y \rangle$$

$$B \to (A \times B)$$

Modalities

$\diamond, \Box$ are semantically vacuous, so only the control rules get a non-trivial interpretation:

$$A \xrightarrow{\lambda x.\langle x, x \rangle} A \times A$$

$$(A \times B) \times C \xrightarrow{\lambda \langle x, y, z \rangle.\langle y, x, z \rangle} B \times (A \times C)$$
Lambda term for simple ellipsis

Bob drinks and Alice does too

A proof of

\[(np \otimes np \backslash s) \otimes ((s\backslash s)/s \otimes (np \otimes (\Diamond (np \backslash s) \otimes \Diamond (np\backslash s) \backslash (np \backslash s))))) \rightarrow s\]

gives term

\[\lambda\langle\text{subj}_1, \text{verb}, \text{coord}, \text{subj}_2, \text{verb}^*, \text{aux}\rangle.\text{(coord ((aux verb*) subj}_2\rangle)(\text{verb subj}_1)\]

The movement and contraction give

\[\lambda\langle\text{subj}_1, \text{verb}, \text{coord}, \text{subj}_2, \text{aux}\rangle.\text{(coord ((aux verb) subj}_2\rangle)(\text{verb subj}_1)\]

Plugging in some constants, we get an abstract term

\[(\text{and } ((\text{dt drinks}) \text{ bob}))(\text{drinks alice}) : s\]
Modelling Vectors with Lambdas

Vector: $\lambda i.v_i$ \hspace{1cm} $I \rightarrow R \ (= V)$

Matrix: $\lambda ij.M_{ij}$ \hspace{1cm} $I \rightarrow I \rightarrow R$

$\odot$: $\lambda vui.v_i \cdot u_i$ \hspace{1cm} $V \rightarrow V \rightarrow R$

Vector $\odot$ Vector: $\lambda v.v \odot v$ \hspace{1cm} $V \rightarrow V$

Matrix $^\top$: $\lambda mij.m_{ji}$ \hspace{1cm} $M \rightarrow M$

Matrix $\times 1$ Vector: $\lambda mvi. \sum_j m_{ij} \cdot v_j$ \hspace{1cm} $M \rightarrow V \rightarrow V$

Cube $\times 2$ Vector: $\lambda cvij. \sum_k c_{ijk} \cdot v_k$ \hspace{1cm} $C \rightarrow V \rightarrow M$

(Muskens & Sadrzadeh, 2016)
Classical Semantics for simple ellipsis

Bob drinks and Alice does too

$$(\lambda P. \lambda Q. P \odot Q (\lambda x. x (\lambda v. \text{drinks } \times_1 v) \text{bob}))((\lambda v. (\text{drinks } \times_1 v) \text{alice}))$$

$$\rightarrow_\beta (\lambda P. \lambda Q. P \odot Q ((\lambda v. \text{drinks } \times_1 v) \text{bob}))((\lambda v. (\text{drinks } \times_1 v) \text{alice}))$$

$$\rightarrow_\beta (\lambda P. \lambda Q. P \odot Q (\text{drinks } \times_1 \text{bob}))(\text{drinks } \times_1 \text{alice})$$

$$\rightarrow_\beta (\text{drinks } \times_1 \text{bob}) \odot (\text{drinks } \times_1 \text{alice})$$
Classical Semantics for Ambiguous Ellipsis

Bob loves Bob’s beer and Alice loves Bob’s beer (sloppy)

\((\text{bob} \times_1 \text{loves} \times_2 (\text{bob} \odot \text{beer})) \odot (\text{alice} \times_1 \text{loves} \times_2 (\text{bob} \odot \text{beer}))\)

Bob loves Bob’s beer and Alice loves Alice’s beer (strict)

\((\text{bob} \times_1 \text{loves} \times_2 (\text{bob} \odot \text{beer})) \odot (\text{alice} \times_1 \text{loves} \times_2 (\text{alice} \odot \text{beer}))\)
Conclusion 1: Classical vs. Quantum Entanglement

Developing Frobenius Semantics fits easily in the DisCoCat framework, but fails to give a proper account for more complex examples of ellipsis.

Classical Semantics are more involved and are non-linear, but give a better account of derivational ambiguity.
LET THE DATA SPEAK
EXPERIMENTING WITH VP ELLIPSIS

▶ GS2011 verb disambiguation dataset (200 samples):

\[
\begin{align*}
\text{man draw photograph} & \sim \text{man attract photograph} \\
\text{man draw photograph} & \sim \text{man depict photograph}
\end{align*}
\]

▶ KS2013 similarity dataset (108 samples):

\[
\text{man bites dog} \sim \text{student achieve result}
\]

▶ We extended the above datasets to elliptical phrases (now with 400/416 sentence pairs)

\[
\text{man bites dog and woman does too} \sim \text{student achieve result and boy does too}
\]

▶ Run experiments with several models:

- **Linear**

\[
\overrightarrow{\text{subj}} \star \overrightarrow{\text{verb}} \star \overrightarrow{\text{obj}} \star \overrightarrow{\text{and}} \star \overrightarrow{\text{subj}}^* \star \overrightarrow{\text{does}} \star \overrightarrow{\text{too}}
\]

- **Non-Linear**

\[
\overrightarrow{\text{subj}} \star \overrightarrow{\text{verb}} \star \overrightarrow{\text{obj}} \star \overrightarrow{\text{subj}}^* \star \overrightarrow{\text{verb}} \star \overrightarrow{\text{obj}}
\]

- **Lambda-Based**

\[
T(\overrightarrow{\text{subj}}, \overrightarrow{\text{verb}}, \overrightarrow{\text{obj}}) \star T(\overrightarrow{\text{subj}}^*, \overrightarrow{\text{verb}}, \overrightarrow{\text{obj}})
\]

- **Picture-Based**

\[
T(\overrightarrow{\text{subj}} \star \overrightarrow{\text{subj}}^*, \overrightarrow{\text{verb}}, \overrightarrow{\text{obj}})
\]

where \(\star\) is addition or multiplication, and \(T\) is some attested model for a transitive sentence.
## EXPERIMENTING WITH VP ELLIPSIS: DISAMBIGUATION RESULTS

<table>
<thead>
<tr>
<th>Method</th>
<th>CB</th>
<th>W2V</th>
<th>GloVe</th>
<th>FT</th>
<th>D2V1</th>
<th>D2V2</th>
<th>ST</th>
<th>IS1</th>
<th>IS2</th>
<th>USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb Only Vector</td>
<td>.4150</td>
<td>.2260</td>
<td>.4281</td>
<td>.2261</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verb Only Tensor</td>
<td>.3039</td>
<td>.4028</td>
<td>.3636</td>
<td>.3548</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add. Linear</td>
<td>.4081</td>
<td>.2619</td>
<td>.3025</td>
<td>.1292</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mult. Linear</td>
<td>.3205</td>
<td>-.0098</td>
<td>.2047</td>
<td>.2834</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add. Non-Linear</td>
<td>.4125</td>
<td>.3130</td>
<td>.3195</td>
<td>.1350</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mult. Non-Linear</td>
<td>.4759</td>
<td>.1959</td>
<td>.2445</td>
<td>.0249</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Lambda</td>
<td>.5078</td>
<td>.4263</td>
<td>.3556</td>
<td>.4543</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Best Lambda</td>
<td>.4949</td>
<td>.4156</td>
<td>.3338</td>
<td>.4278</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Picture</td>
<td>.5080</td>
<td>.4263</td>
<td>.3916</td>
<td>.4572</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Method                        |      |     |       |     |     |      |      |      |       |      |
| Sent Encoder                  | .1425 | .2369 | -.1764 | .3382 | .3477 | .2564 |
| Sent Encoder+Res              | .2269 | .3021 | -.1607 | .3437 | .3129 | .2576 |
| Sent Encoder-Log              | .1840 | .2500 | -.1252 | **.3484** | .3241 | .2252 |

**Table**: Spearman $\rho$ scores for the ellipsis disambiguation experiment. **CB**: count-based, **W2V**: Word2Vec, **FT**: FastText, **ST**: Skip-Thoughts, **IS1**: InferSent (GloVe), **IS2**: InferSent (FastText), **USE**: Universal Sentence Encoder.
## EXPERIMENTING WITH VP ELLIPSIS: SIMILARITY RESULTS

<table>
<thead>
<tr>
<th>Method</th>
<th>CB</th>
<th>W2V</th>
<th>GloVe</th>
<th>FT</th>
<th>D2V1</th>
<th>D2V2</th>
<th>ST</th>
<th>IS1</th>
<th>IS2</th>
<th>USE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb Only Vector</td>
<td>0.4562</td>
<td>0.5833</td>
<td>0.4348</td>
<td>0.6513</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verb Only Tensor</td>
<td>0.3946</td>
<td>0.5664</td>
<td>0.4426</td>
<td>0.5337</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add. Linear</td>
<td>0.7000</td>
<td>0.7258</td>
<td>0.6964</td>
<td>0.7408</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mult. Linear</td>
<td>0.6330</td>
<td>0.1302</td>
<td>0.3666</td>
<td>0.1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add. Non-Linear</td>
<td>0.6808</td>
<td>0.7617</td>
<td>0.7103</td>
<td>0.7387</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mult. Non-Linear</td>
<td>0.7237</td>
<td>0.3550</td>
<td>0.2439</td>
<td>0.4500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Lambda</td>
<td>0.7410</td>
<td>0.7061</td>
<td>0.4907</td>
<td>0.6989</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Best Lambda</td>
<td>0.7370</td>
<td>0.6713</td>
<td>0.4819</td>
<td>0.6871</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best Picture</td>
<td>0.7413</td>
<td>0.7105</td>
<td>0.4907</td>
<td>0.7085</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sent Encoder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.5901</td>
<td>0.6188</td>
<td>0.5851</td>
<td>0.7785</td>
<td>0.7009</td>
<td>0.6463</td>
</tr>
<tr>
<td>Sent Encoder+Res</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.6878</td>
<td>0.6875</td>
<td>0.6039</td>
<td>0.8022</td>
<td>0.7486</td>
<td>0.6791</td>
</tr>
<tr>
<td>Sent Encoder-Log</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.1840</td>
<td>0.6599</td>
<td>0.4715</td>
<td>0.7815</td>
<td>0.7301</td>
<td>0.6397</td>
</tr>
</tbody>
</table>

**Table:** Spearman $\rho$ scores for the ellipsis similarity experiment. **CB:** count-based, **W2V:** Word2Vec, **FT:** FastText, **ST:** Skip-Thoughts, **IS1:** InferSent (GloVe), **IS2:** InferSent (FastText), **USE:** Universal Sentence Encoder.
Conclusion 2: Classical vs. Quantum Entanglement

Experimentally, the linear approximation that Frobenius Semantics gives is equally performant to the classical semantics!
Future Work

1. Entailment:
   
   Dogs sleep and cats too ⇒ ⌈ cats walk

2. Guess the antecedent (ambiguity!):
   
   Dogs run, cats walk, and foxes ...

3. Negation:
   
   Dogs sleep but cats do not.
Thank you!

g.j.wijnholds@qmul.ac.uk