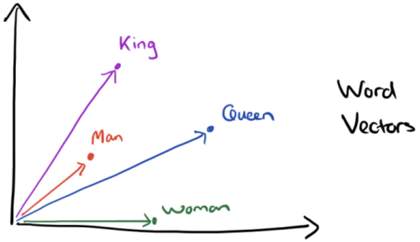
Classical Copying versus Quantum Entanglement in Natural Language: the Case of VP-ellipsis

Gijs Jasper Wijnholds[1]   Mehrnoosh Sadrzadeh[1]

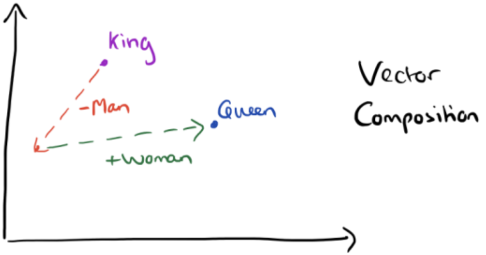
Queen Mary University of London, United Kingdom
g.j.wijnholds@qmul.ac.uk

SYCO 2
December 17, 2018

DISTRIBUTIONAL SEMANTICS: MEANING IN CONTEXT



COMPOSING WORD EMBEDDINGS: A CHALLENGE



	Coordination	$\overrightarrow{\textbf{dancing and running}} = ??$
	Quantification	$\overrightarrow{\textbf{Every student likes some teacher}} = ??$
	Anaphora	$\overrightarrow{\textbf{shaves himself}} = ??$
$\Rightarrow$	**Ellipsis**	$\overrightarrow{\textbf{Matt went to Croatia and Max did too}} = ??$

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.

$$\text{Bob } \underbrace{\text{drinks a beer}}_{\text{ant VP}} \text{ and Alice } \underbrace{\text{does too}}_{\text{marker}}$$

ELLIPSIS NEEDS COPYING AND MOVEMENT

Bob	drinks a beer	and	Alice	␣	does too
	ANT VP				MARKER

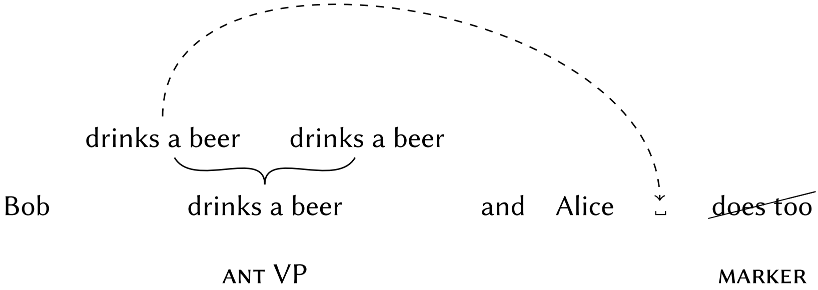
ELLIPSIS NEEDS COPYING AND MOVEMENT

drinks a beer   drinks a beer

Bob   drinks a beer   and   Alice   ␣   does too

ANT VP   MARKER

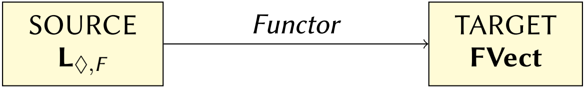
ELLIPSIS NEEDS COPYING AND MOVEMENT



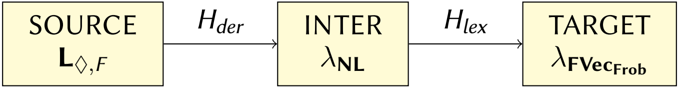
THE CHALLENGE: COMPOSE WORD VECTORS TO GET A MEANING REPRESENTATION FOR VP ELLIPSIS

THE BIG PICTURE

Quantum Entanglement



Classical



QUANTUM ENTANGLEMENT

LAMBEK VS. LAMBEK

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

$$B \otimes B\backslash A \to A \qquad A \to B\backslash(B \otimes A)$$

$$A/B \otimes B \to A \qquad A \to (A \otimes B)/B$$

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

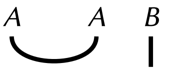
Interpretation: words have types, and type-respecting embeddings

Word	**Type**	**embedding**
john	np	$\overrightarrow{john} \in N$
sleeps	$np\backslash s$	$\overline{sleep} \in N \otimes S$ ($\leftarrow$ matrix)
likes	$(np\backslash s)/np$	$\overline{like} \in N \otimes S \otimes N$ ($\leftarrow$ cube)
beer	np	$\overrightarrow{beer} \in N$

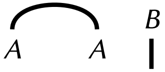
(Coecke et al., 2013)

IN PICTURES

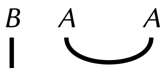
$A \otimes A\backslash B \rightarrow B$



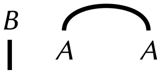
$B \rightarrow A\backslash(A \otimes B)$



$B/A \otimes A \rightarrow B$



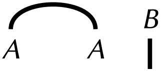
$B \rightarrow (B \otimes A)/A$



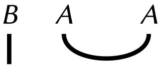
(Coecke et al., 2013)

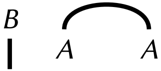
IN PICTURES

$A \otimes A \backslash B \to B$ $\qquad$ $B \to A \backslash (A \otimes B)$



$B / A \otimes A \to B$ $\qquad$ $B \to (B \otimes A) / A$





LINEAR!!

(Coecke et al., 2013)

LAMBEK WITH CONTROL OPERATORS: $\mathbf{L}_{\Diamond,F}$

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

$$B \otimes B\backslash A \rightarrow A \qquad A \rightarrow B\backslash(B \otimes A)$$

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

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

Modalities: application, co-application

$$\Diamond\Box A \rightarrow A \qquad A \rightarrow \Box\Diamond A$$

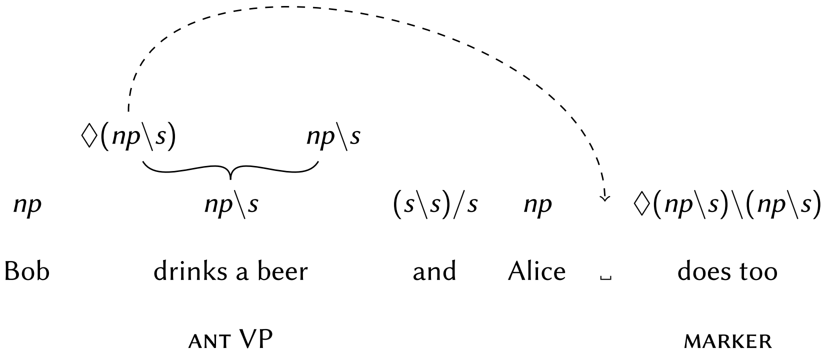
- Linear logic: controlled duplication/deletion of resources via $! = \Diamond\Box$. Here: controlled copying, reordering

Controlled contraction, commutativity

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

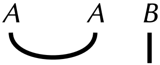
$$\Diamond A \otimes (\Diamond B \otimes C) \rightarrow \Diamond B \otimes (\Diamond A \otimes C)$$

ILLUSTRATION

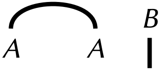


IN PICTURES

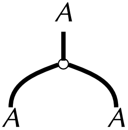
$A \otimes A\backslash B \to B$



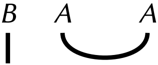
$B \to A\backslash(A \otimes B)$



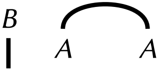
$A \to \Diamond A \otimes A$



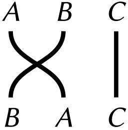
$B/A \otimes A \to B$



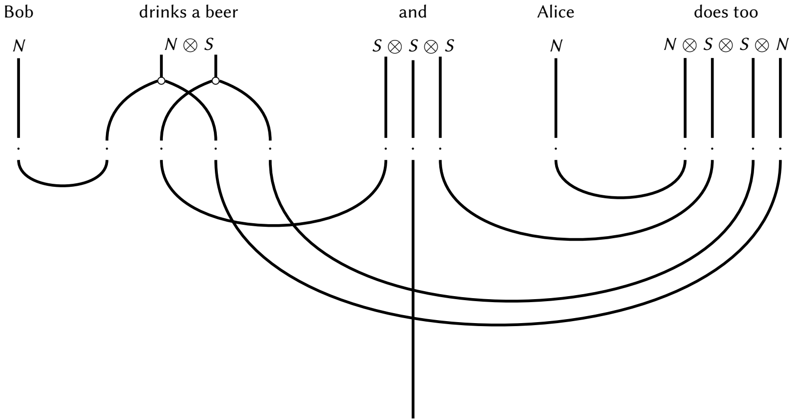
$B \to (B \otimes A)/A$



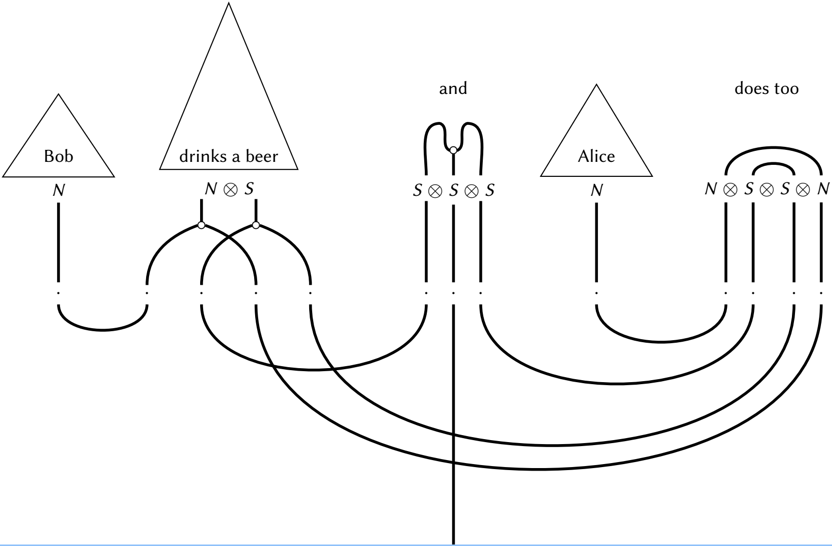
$(\Diamond A \otimes B) \otimes C \to B \otimes (\Diamond A \otimes C)$



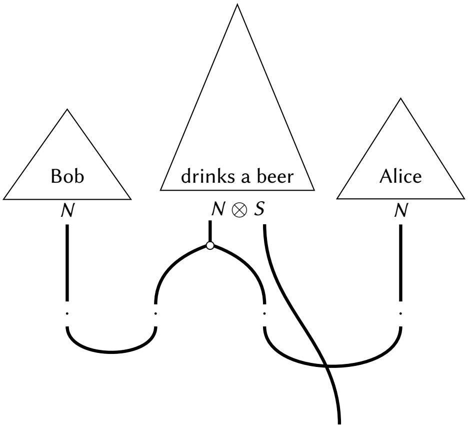
QUANTUM ENTANGLEMENT AND ELLIPSIS



QUANTUM ENTANGLEMENT AND ELLIPSIS



QUANTUM ENTANGLEMENT AND ELLIPSIS



$(\overrightarrow{Bob} \odot \overrightarrow{Alice})^\top \times \overrightarrow{drinks\ a\ beer}$

A MORE COMPLICATED CASE: SLOPPY READING

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

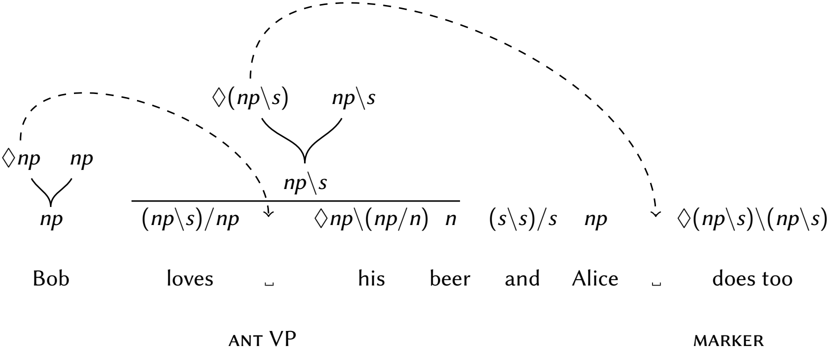
Bob loves Bob's beer and Alice loves Bob's beer

np	$(np\backslash s)/np$		$\Diamond np\backslash(np/n)$	n	$(s\backslash s)/s$	np		$\Diamond(np\backslash s)\backslash(np\backslash s)$
Bob	loves	␣	his	beer	and	Alice	␣	does too
		ANT VP						MARKER

A MORE COMPLICATED CASE: SLOPPY READING

Bob loves his beer and Alice does too ⇝

Bob loves Bob's beer and Alice loves Bob's beer



A MORE COMPLICATED CASE: STRICT READING

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

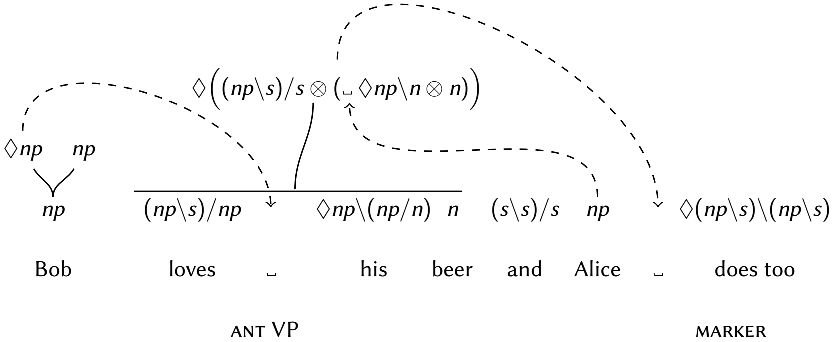
Bob loves Bob's beer and Alice loves Alice's beer

np	$(np\backslash s)/np$		$\Diamond np\backslash(np/n)$	n	$(s\backslash s)/s$	np		$\Diamond(np\backslash s)\backslash(np\backslash s)$
Bob	loves	␣	his	beer	and	Alice	␣	does too
		ANT VP						MARKER

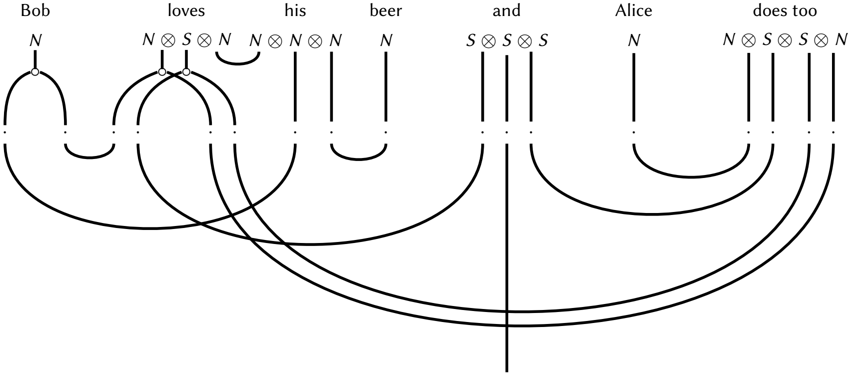
A MORE COMPLICATED CASE: STRICT READING

Bob loves his beer and Alice does too ⇝

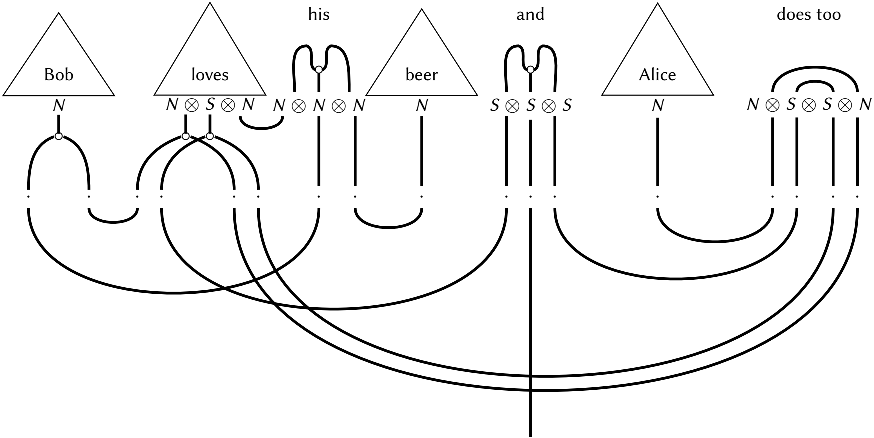
Bob loves Bob's beer and Alice loves Alice's beer



A More Complicated Case: Sloppy Reading

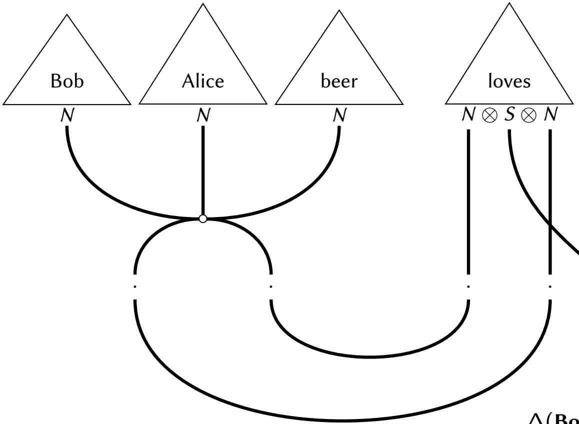


A More Complicated Case: Sloppy Reading



A More Complicated Case: Sloppy Reading

Bob loves Bob's beer and Alice loves Bob's beer



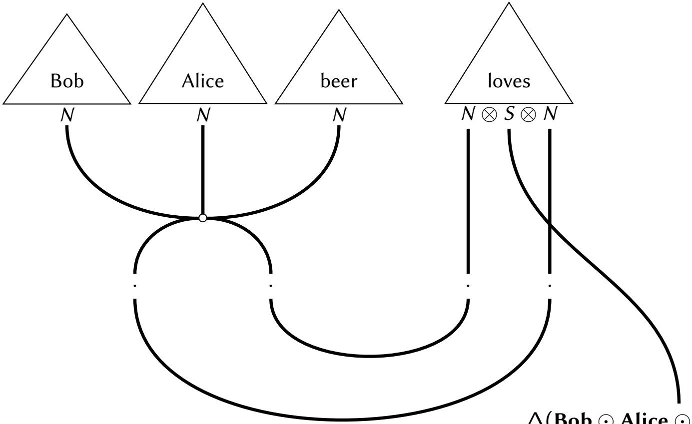
$\Delta(\mathbf{Bob} \odot \mathbf{Alice} \odot \mathbf{Beer})_{ik}\mathbf{loves}_{ijk}$

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



$\Delta(\mathbf{Bob} \odot \mathbf{Alice} \odot \mathbf{Beer})_{ik}\mathbf{loves}_{ijk}$

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

$$\lfloor A \otimes B \rfloor = \lfloor A \rfloor \times \lfloor B \rfloor \quad \lfloor A/B \rfloor = \lfloor A \rfloor \to \lfloor B \rfloor \quad \lfloor A \backslash B \rfloor = \lfloor A \rfloor \to \lfloor B \rfloor \quad \lfloor \Diamond A \rfloor = \lfloor \Box A \rfloor = \lfloor A \rfloor$$

Application, co-application

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

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

Modalities

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

$$A \xrightarrow{\lambda x.\langle x, x\rangle} A \times A \qquad\qquad (A \times B) \times C \xrightarrow{\lambda\langle x, y, z\rangle.\langle y, x, z\rangle} B \times (A \times C)$$

Lambda term for simple ellipsis

Bob drinks and Alice does too

A proof of

$$(np \otimes np\backslash s) \otimes ((s\backslash s)/s \otimes (np \otimes (\Diamond(np\backslash s) \otimes \Diamond(np\backslash s)\backslash(np\backslash s)))) \longrightarrow s$$

gives term

$$\lambda\langle \mathtt{subj}_1, \mathtt{verb}, \mathtt{coord}, \mathtt{subj}_2, \mathtt{verb}^*, \mathtt{aux}\rangle.(\mathtt{coord}\,((\mathtt{aux}\,\mathtt{verb}^*)\,\mathtt{subj}_2))(\mathtt{verb}\,\mathtt{subj}_1)$$

The movement and contraction give

$$\lambda\langle \mathtt{subj}_1, \mathtt{verb}, \mathtt{coord}, \mathtt{subj}_2, \mathtt{aux}\rangle.(\mathtt{coord}\,((\mathtt{aux}\,\mathtt{verb})\,\mathtt{subj}_2))(\mathtt{verb}\,\mathtt{subj}_1)$$

Plugging in some constants, we get an abstract term

$$(\mathtt{and}\,((\mathtt{dt}\,\mathtt{drinks})\,\mathtt{bob}))(\mathtt{drinks}\,\mathtt{alice}) : s$$

Modelling Vectors with Lambdas

Vector:	$\lambda i.v_i$	$I \to R \quad (= V)$
Matrix:	$\lambda ij.M_{ij}$	$I \to I \to R$
$\odot$:	$\lambda vui.v_i \cdot u_i$	$V \to V \to R$
Vector $\odot$ **Vector:**	$\lambda v.v \odot v$	$V \to V$
Matrix$^\top$	$\lambda mij.m_{ji}$	$M \to M$
Matrix $\times_1$ **Vector**	$\lambda mvi.\sum_j m_{ij} \cdot v_j$	$M \to V \to V$
Cube $\times_2$ **Vector**	$\lambda cvij.\sum_k c_{ijk} \cdot v_k$	$C \to V \to M$

(Muskens & Sadrzadeh, 2016)

Classical Semantics for simple ellipsis

Bob drinks and Alice does too

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

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

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

$$\rightarrow_\beta\ (\mathbf{drinks} \times_1 \mathbf{bob}) \odot (\mathbf{drinks} \times_1 \mathbf{alice})$$

Classical Semantics for Ambiguous Ellipsis

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

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

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

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

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):

$$\underline{\text{man draw photograph}} \sim \underline{\text{man attract photograph}}$$
$$\underline{\text{man draw photograph}} \sim \underline{\text{man depict photograph}}$$

- KS2013 similarity dataset (108 samples):

$$\underline{\text{man bites dog}} \sim \underline{\text{student achieve result}}$$

- We extended the above datasets to elliptical phrases (now with 400/416 sentence pairs)
$\underline{\text{man bites dog and woman does too}} \sim \underline{\text{student achieve result and boy does too}}$
- Run experiments with several models:

Linear	$\overrightarrow{subj} \star \overrightarrow{verb} \star \overrightarrow{obj} \star \overrightarrow{and} \star \overrightarrow{subj^*} \star \overrightarrow{does} \star \overrightarrow{too}$
Non-Linear	$\overrightarrow{subj} \star \overrightarrow{verb} \star \overrightarrow{obj} \star \overrightarrow{subj^*} \star \overrightarrow{verb} \star \overrightarrow{obj}$
Lambda-Based	$T(\overrightarrow{subj}, \overline{verb}, \overrightarrow{obj}) \star T(\overrightarrow{subj^*}, \overline{verb}, \overrightarrow{obj})$
Picture-Based	$T(\overrightarrow{subj} \star \overrightarrow{subj^*}, \overline{verb}, \overrightarrow{obj})$

where $\star$ is addition or multiplication, and T is some attested model for a transitive sentence.

EXPERIMENTING WITH VP ELLIPSIS: DISAMBIGUATION RESULTS

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

	CB	W2V	GloVe	FT	D2V1	D2V2	ST	IS1	IS2	USE
Verb Only Vector	.4562	.5833	.4348	.6513						
Verb Only Tensor	.3946	.5664	.4426	.5337						
Add. Linear	.7000	.7258	.6964	.7408						
Mult. Linear	.6330	.1302	.3666	.1995						
Add. Non-Linear	.6808	**.7617**	.7103	.7387						
Mult. Non-Linear	.7237	.3550	.2439	.4500						
Best Lambda	.7410	.7061	.4907	.6989						
2nd Best Lambda	.7370	.6713	.4819	.6871						
Best Picture	.7413	.7105	.4907	.7085						
Sent Encoder					.5901	.6188	.5851	.7785	.7009	.6463
Sent Encoder+Res					.6878	.6875	.6039	**.8022**	.7486	.6791
Sent Encoder-Log					.1840	.6599	.4715	.7815	.7301	.6397

Table: Spearman ρ 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 $\Rightarrow \lightning$ 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