5th International Conference on the
History and Philosophy of Computing (HaPoC)
28-30 Oct 2019 Bergamo (Italy)

The Logic of Language

An alternative logical approach to machine learning based on the case of Natural Language Processing

Juan Luis Gastaldi
HPM – ETH Zürich
juan.luis.gastaldi@gess.ethz.ch

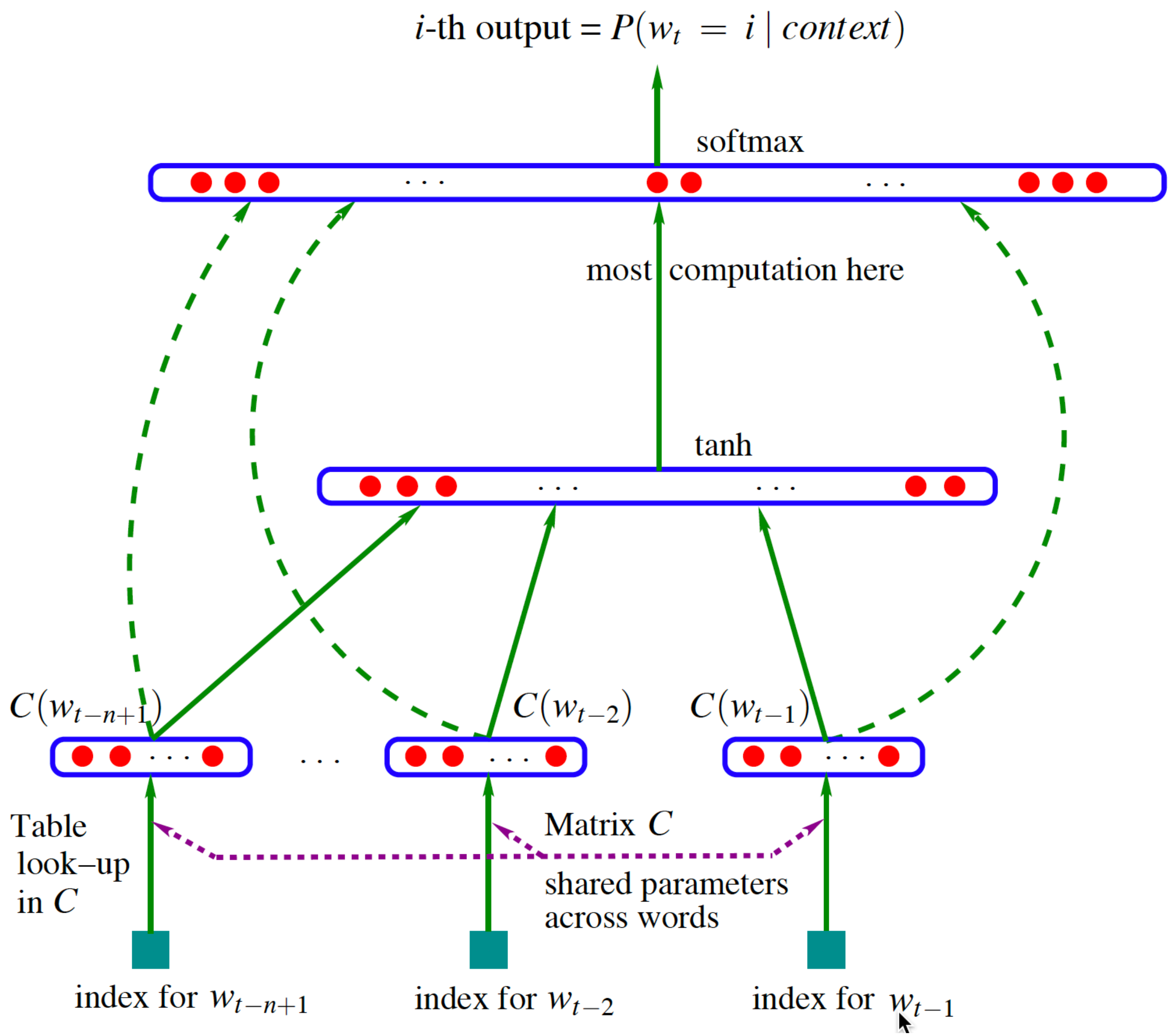
Luc Pellissier
IRIF – Paris Diderot
pellissier@irif.fr

This project has received funding from the
European Union's Horizon 2020 research and innovation programme
under grant agreement No 839730

- WORD EMBEDDINGS
- THEORETICAL INSIGHTS
- ABSTRACT MACHINES
- BIORTHOGONAL TYPING
- SOME PROOFS OF CONCEPT
- CONCLUSIONS

Index	Word
...	...
535	nearly
536	shares
537	member
538	campaign
539	media
540	needs
541	why
542	house
543	issues
544	costs
545	fire
...	...

$$v_{house} = (\underbrace{0,0,0,0,0,0,0,0,\ldots,0,\overbrace{1}^{542^{\text{nd}}\text{ position}},0,\ldots,0,0,0,0,0,0,0,0}_{3\text{ million dimensions}})$$



Source: Bengio et al., 2003

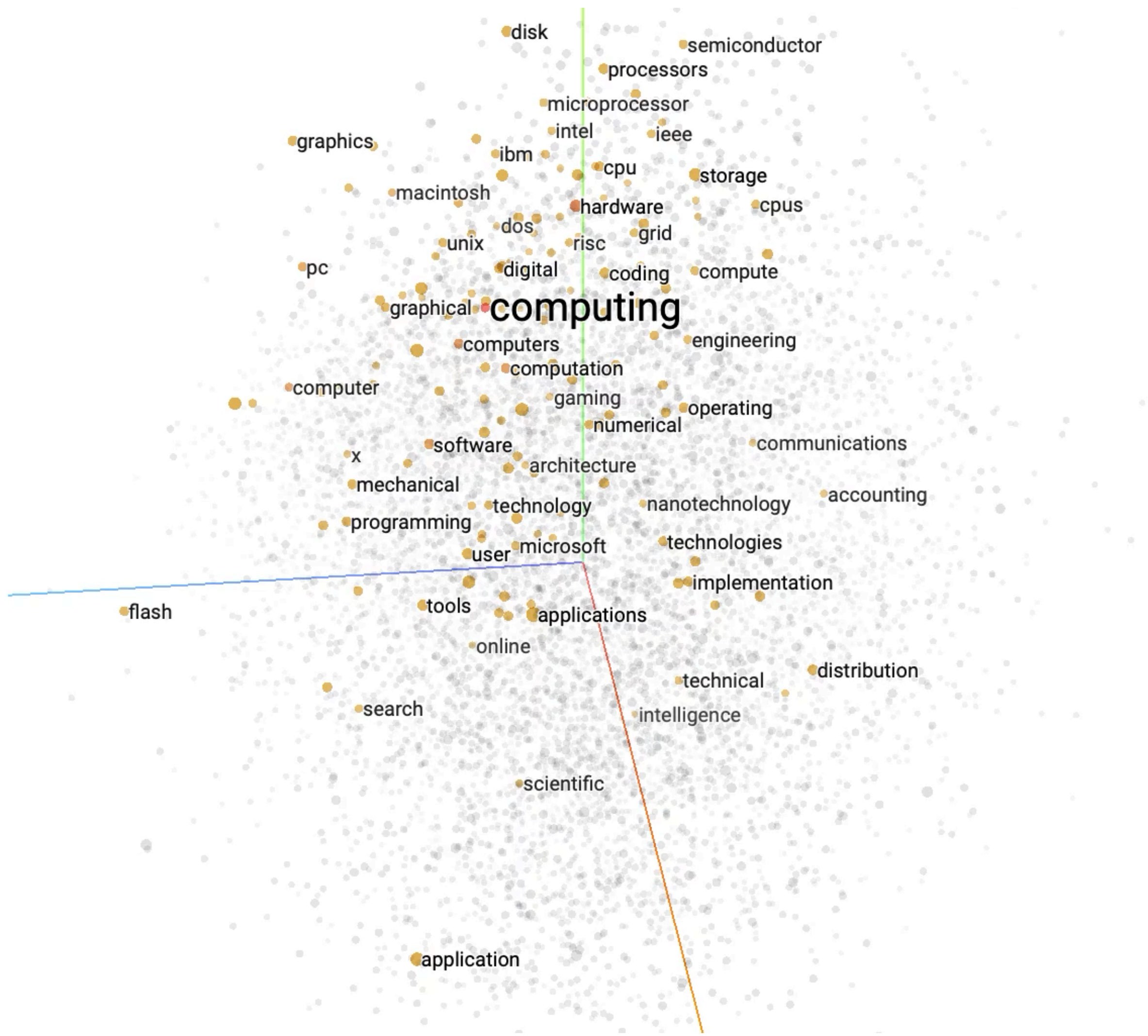
$$v_{house} = (\underbrace{0,0,0,0,0,0,0,0,\ldots,0,\overbrace{1}^{542^{\text{nd}}\text{ position}},0,\ldots,0,0,0,0,0,0,0,0}_{\text{3 million dimensions}})$$

$$v_{house} = (\underbrace{0, 0, 0, 0, 0, 0, 0, 0, \ldots, 0, \overbrace{1}^{542^{\mathrm{nd}}\ \text{position}}, 0, \ldots, 0, 0, 0, 0, 0, 0, 0, 0}_{3\ \text{million dimensions}})$$

$$v_{house} = (\underbrace{0.157227, -0.0708008, 0.0539551, \ldots, -0.041748, 0.00982666, -0.00494385, -0.032959}_{300\ \text{dimensions}})$$

house	cosine distance*
houses	0.292761
bungalow	0.312144
apartment	0.3371
bedroom	0.350306
townhouse	0.361592
residence	0.380158
mansion	0.394181
farmhouse	0.414243
duplex	0.424206
homes	0.43802

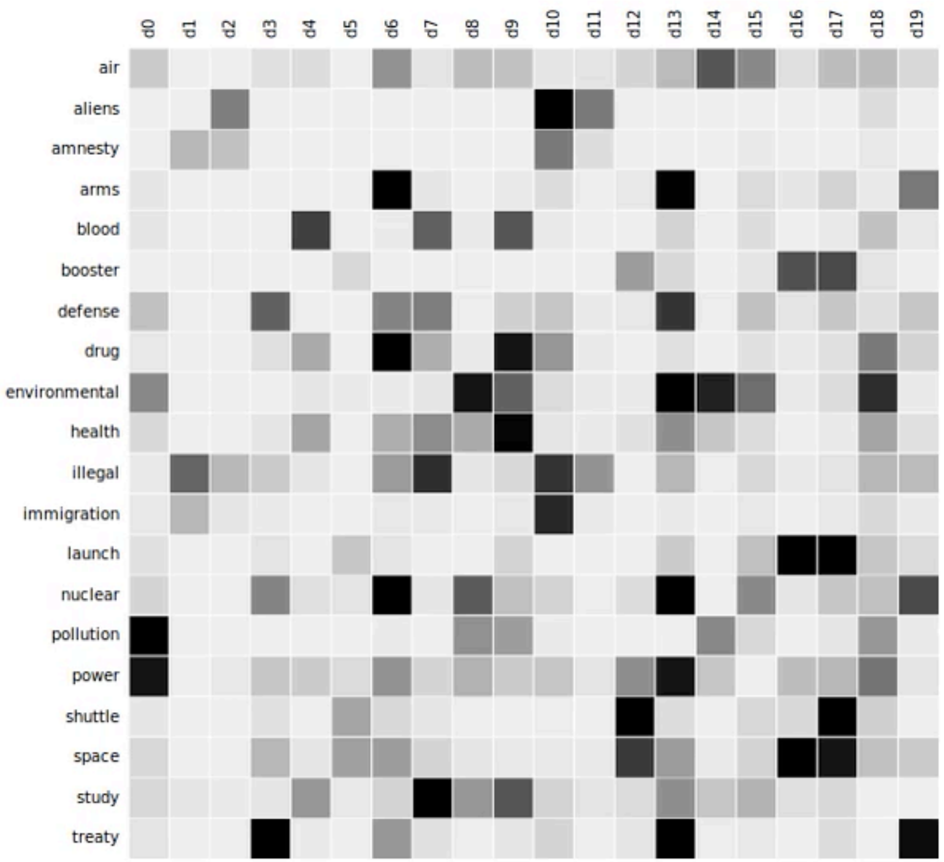
$$* \, cosdistance(u, v) = 1 - \frac{u \cdot v}{\|u\|\|v\|}$$


disk
semiconductor
processors
microprocessor
intel
ieee
graphics
ibm
cpu
storage
macintosh
hardware
cpus
dos
unix
risc
grid
pc
digital
coding
compute
graphical
computing
computers
engineering
computation
computer
gaming
operating
numerical
communications
software
x
architecture
mechanical
nanotechnology
accounting
technology
programming
microsoft
technologies
user
implementation
flash
tools
applications
online
distribution
technical
search
intelligence
scientific
application

- I like deep learning.
- I like NLP.
- I enjoy flying.

counts	I	like	enjoy	deep	learning	NLP	flying	.
I	0	2	1	0	0	0	0	0
like	2	0	0	1	0	1	0	0
enjoy	1	0	0	0	0	0	1	0
deep	0	1	0	0	1	0	0	0
learning	0	0	0	1	0	0	0	1
NLP	0	1	0	0	0	0	0	1
flying	0	0	1	0	0	0	0	1
.	0	0	0	0	1	1	1	0

Source: Manning & Socher, Stanford CS224n course, 2017

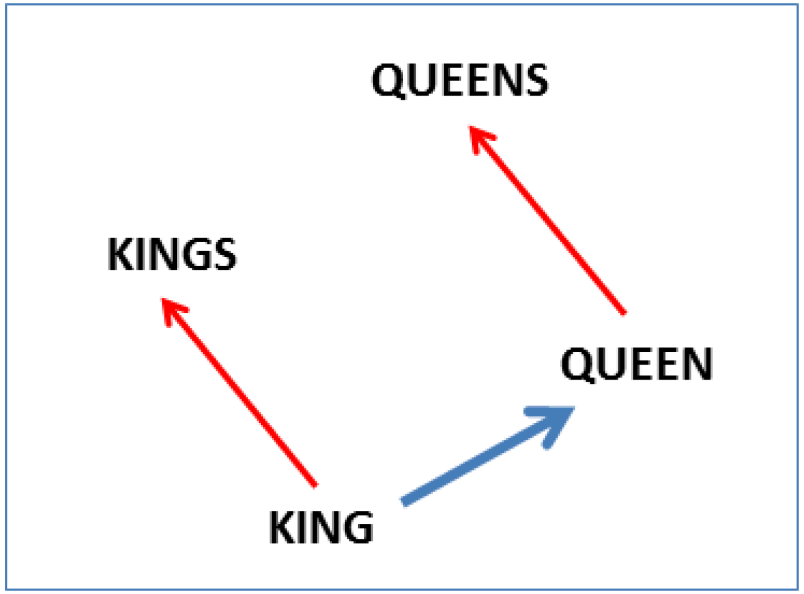


THE EMBEDDING SPACE IS STRUCTURED

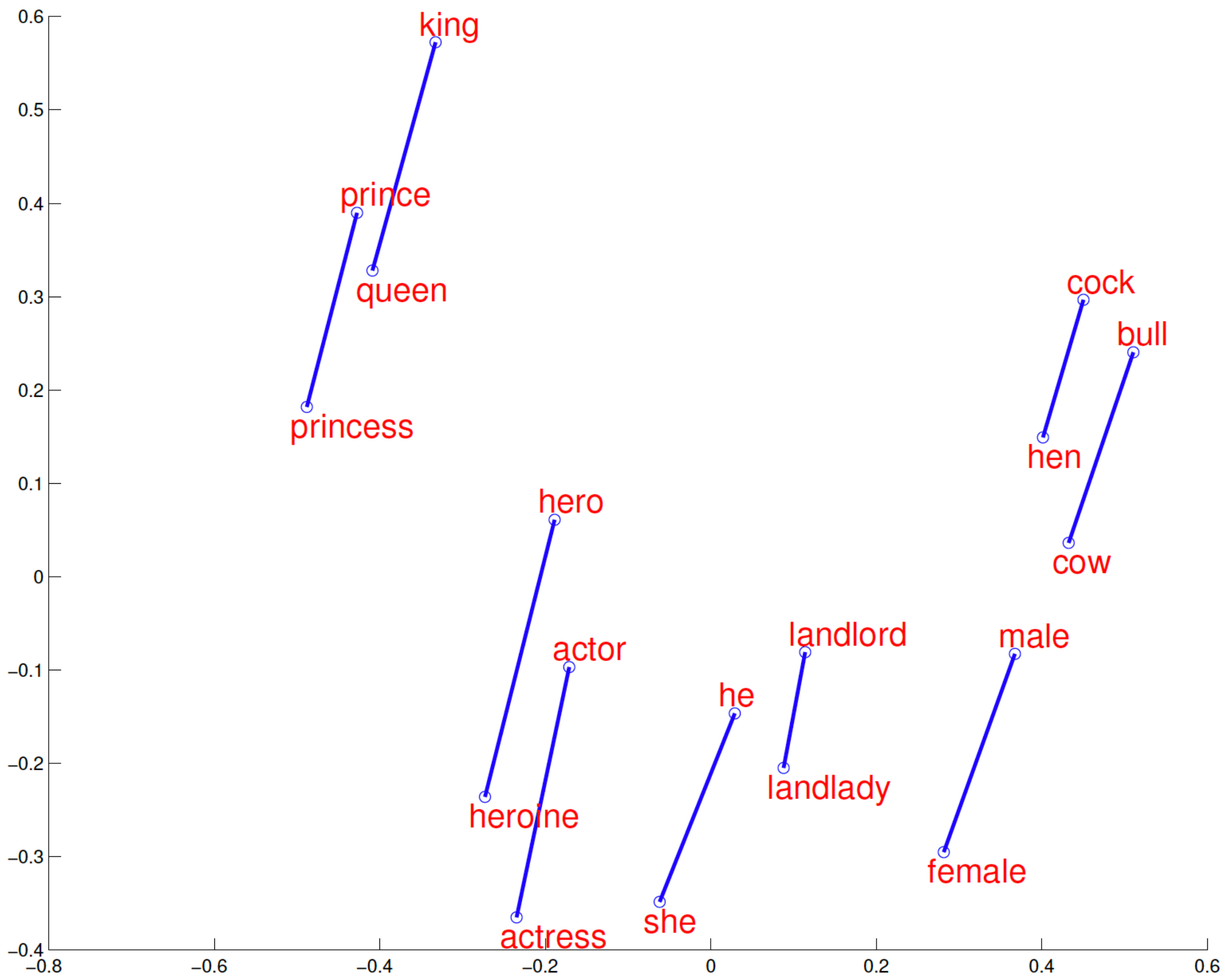
$$v_{house} - v_{city} + v_{countryside} \approx v_{farmhouse}$$

$$v_{king} - v_{man} + v_{woman} \approx v_{queen}$$

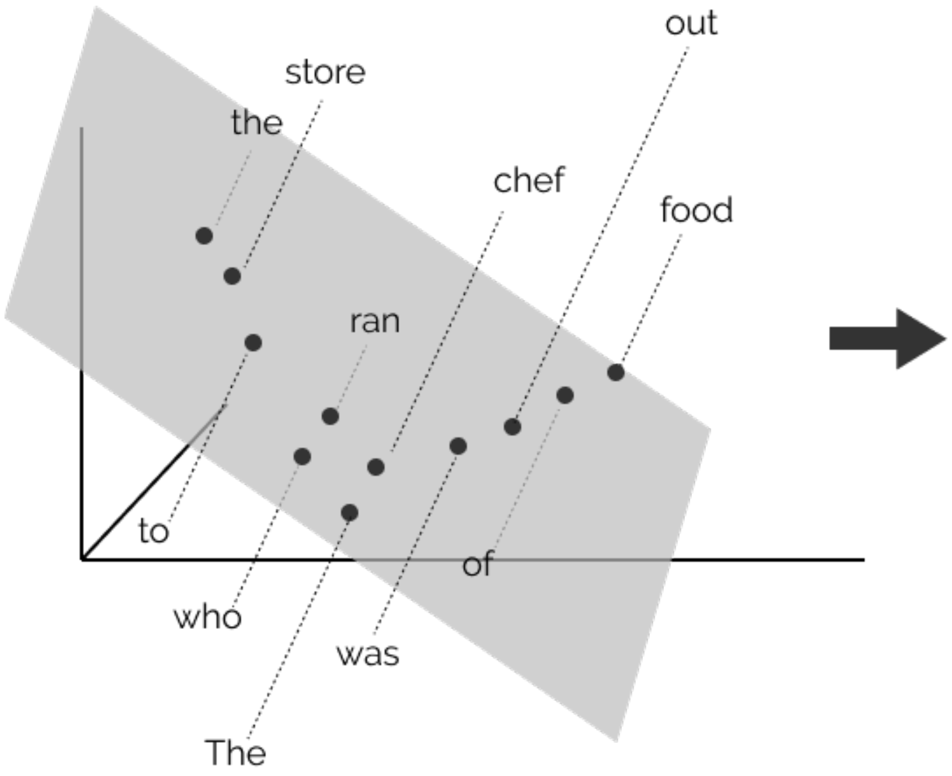
$$v_{king} - v_{queen} \approx v_{man} - v_{woman}$$

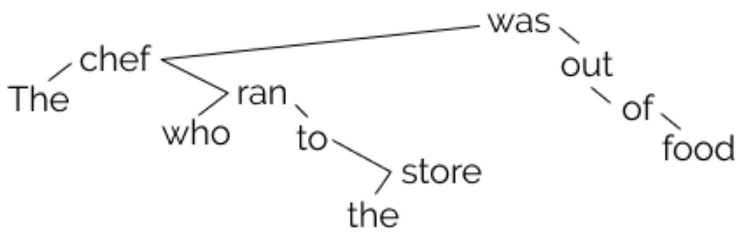


Source: Mikolov et al., 2013.



Source: Mikolov et al., 2013.





Source: Hewit et al., 2019.

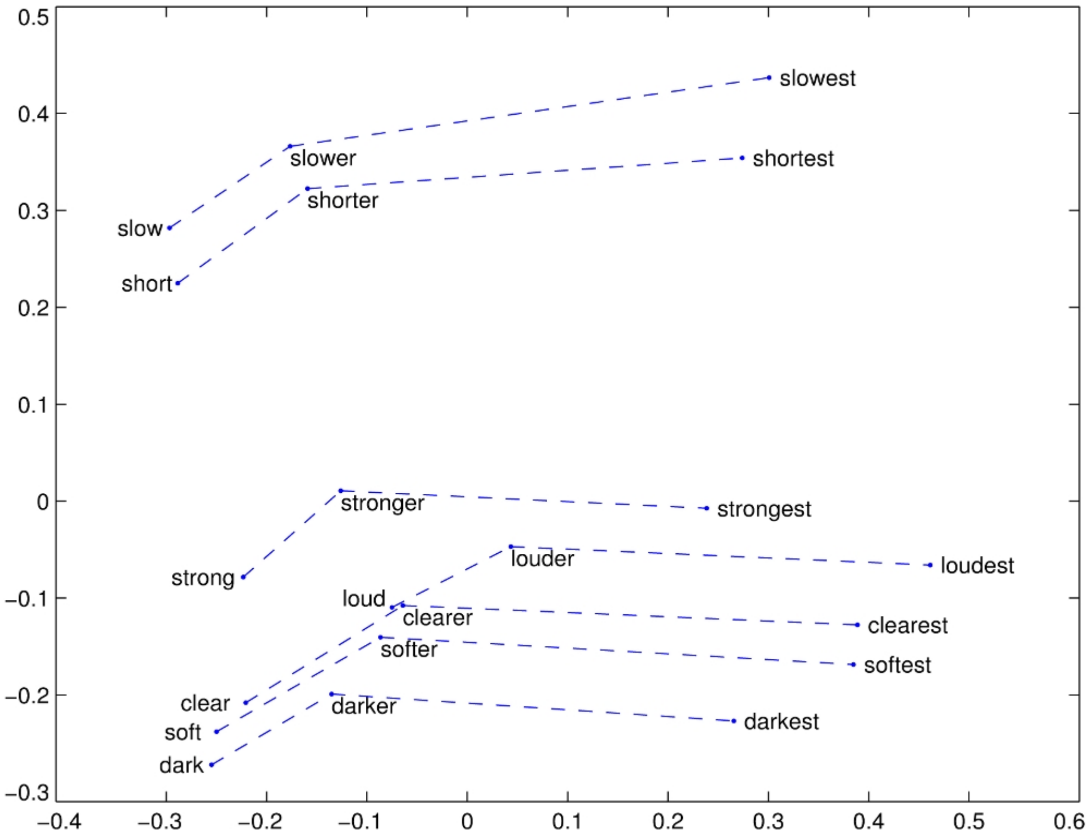
Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

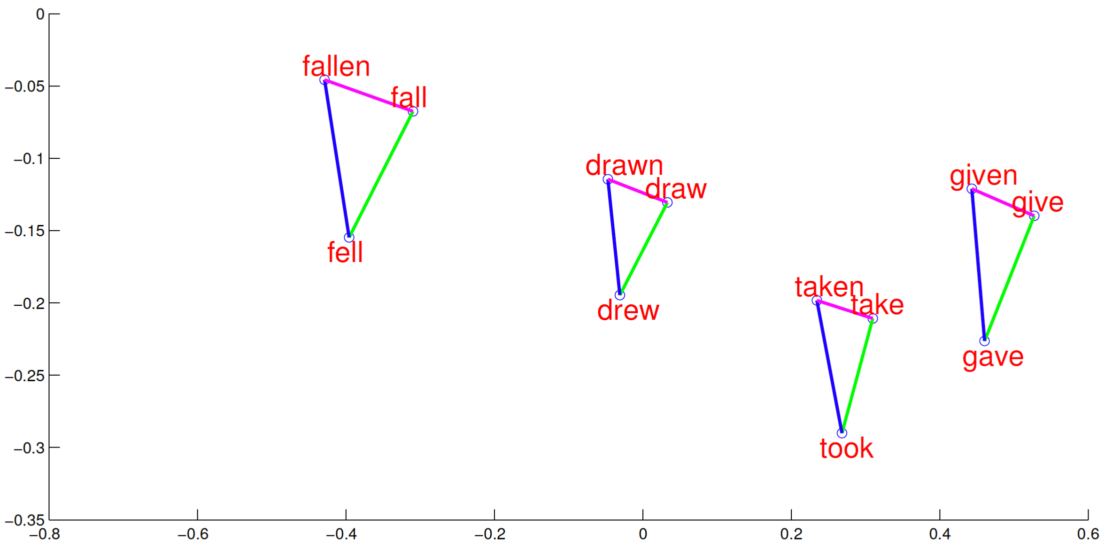
Source: Mikolov et al., 2013.

Syntax and Semantics Continuum

Type of relationship	Word Pair 1		Word Pair 2	
Common capital city	Athens	Greece	Oslo	Norway
All capital cities	Astana	Kazakhstan	Harare	Zimbabwe
Currency	Angola	kwanza	Iran	rial
City-in-state	Chicago	Illinois	Stockton	California
Man-Woman	brother	sister	grandson	granddaughter
Adjective to adverb	apparent	apparently	rapid	rapidly
Opposite	possibly	impossibly	ethical	unethical
Comparative	great	greater	tough	tougher
Superlative	easy	easiest	lucky	luckiest
Present Participle	think	thinking	read	reading
Nationality adjective	Switzerland	Swiss	Cambodia	Cambodian
Past tense	walking	walked	swimming	swam
Plural nouns	mouse	mice	dollar	dollars
Plural verbs	work	works	speak	speaks

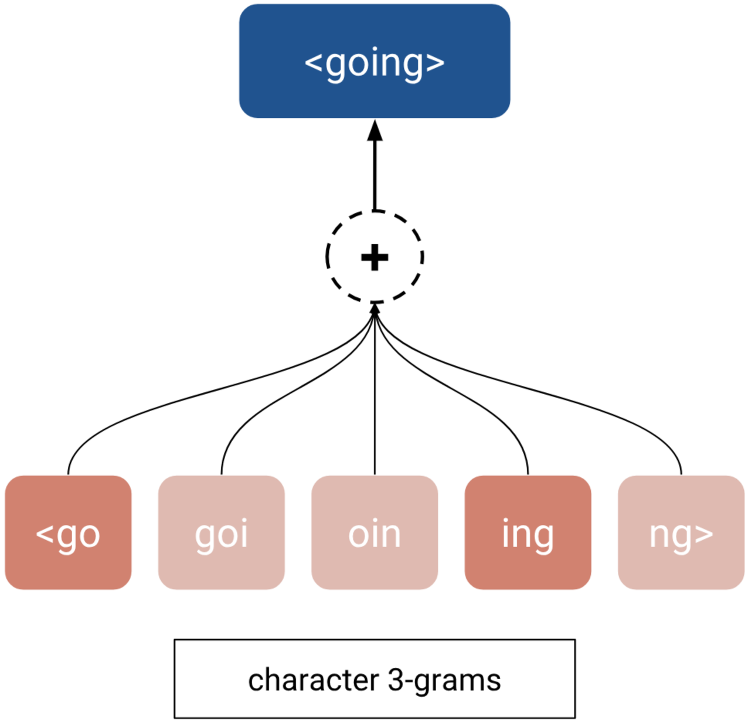
Source: Mikolov et al., 2013.



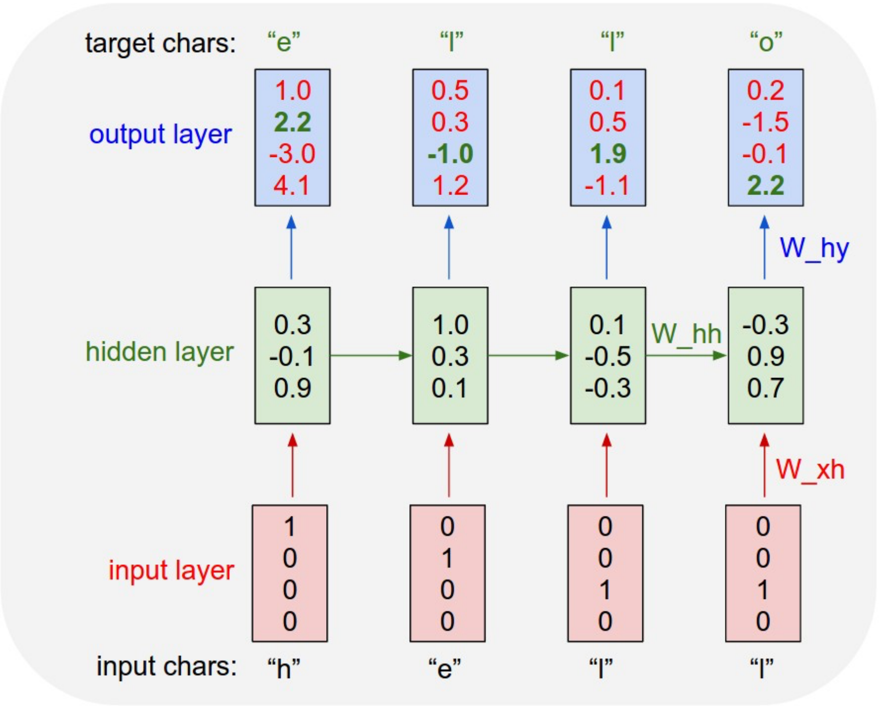


Source: Pennington et al., 2014.

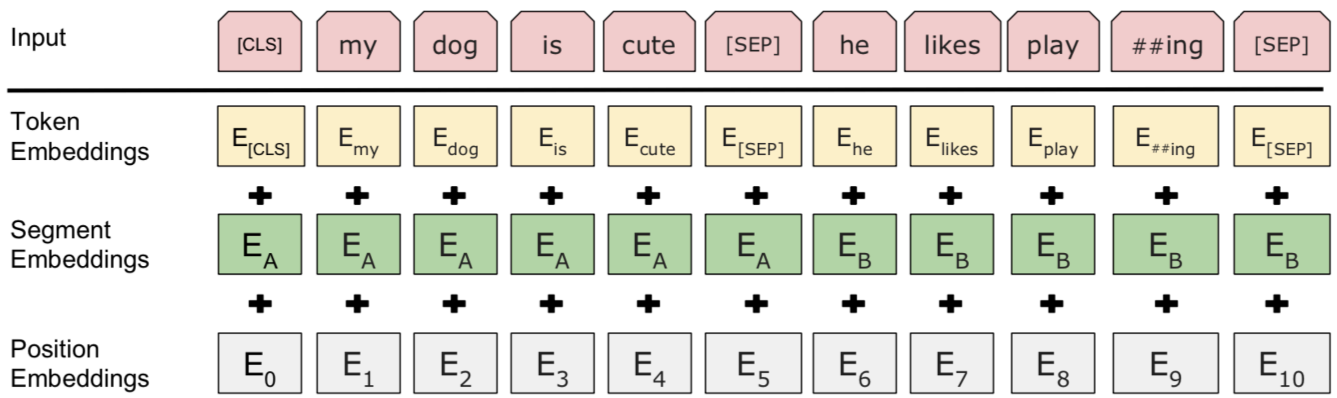
Source: Mikolov et al., 2013.



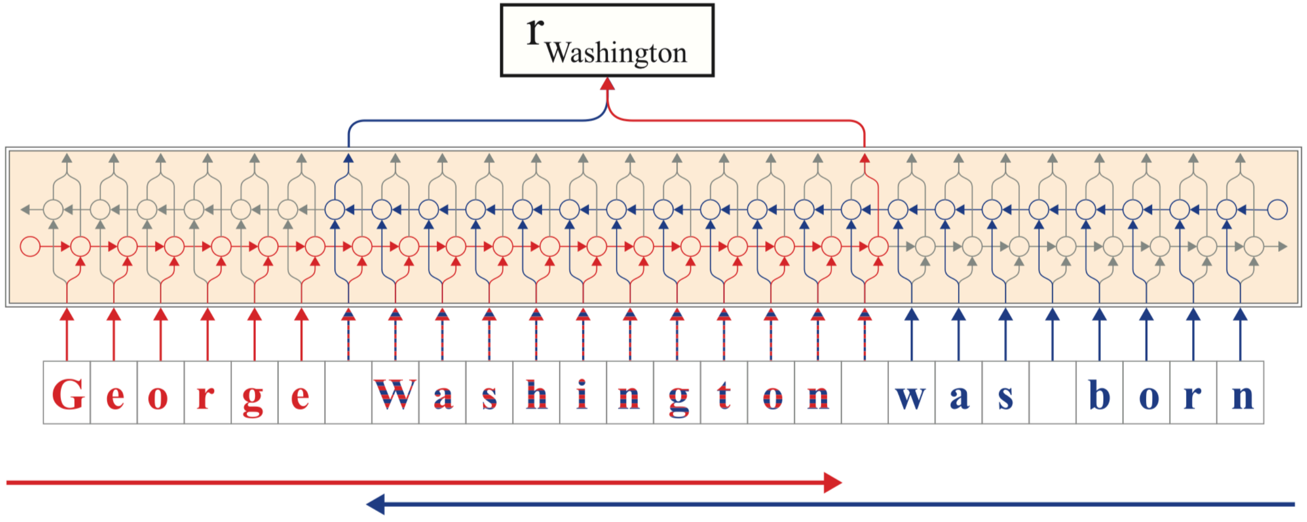
Source: El Boukkouri, 2018



Source: Karpathy, 2015



Source: Deblin et al., 2018



Source: Akbik et al., 2018

- The embedding space is not only organized in neighborhoods but highly structured following relatively precise dimensions

- Syntax and semantics are not two clearly distinguished dimensions of language, but rather zones of a continuum resulting from one and the same analytic procedure

- Words are not the fundamental units of language; language is made of interrelated units at different levels, including non significant ones (letters)

- If DNNs models are above all classificatory devices, their relation to natural language shows that classification is not about mutually exclusive classes resulting from equivalence relations over terms, but a deeper structure emerging from the interaction between terms at different levels

- Those properties are independent from the specific technique of DNNs

Computation

Models of computation

Turing machines operational, mechanical

λ-calculus functional

$$(\lambda x \cdot t)u \to t\,[x \leftarrow u]$$

Implementation of the λ-calculus

Translate the high-level concepts to low-level ones

- non-deterministic (but still confluent)
- no distinction between data and operation

$\rightsquigarrow$ introduce a middle ground

The Krivine Abstract Machine

Specifies the order of evaluation by focusing on part of a term.
Two components: a term and a context: a sequence of terms $t_1 \bullet (t_2 \bullet \cdots)$.

Two rules

$$\begin{array}{r|l c r|l r} tu & \pi & \longrightarrow & t & u \bullet \pi & \text{(push)} \\ \lambda x{\cdot}t & u \bullet \pi & \longrightarrow & t\,[x \leftarrow u] & \pi & \text{(pop)} \end{array}$$

Allows to simulate the λ-calculus.

- Deterministic
- active data: the term
- inactive data: the context
- two operations

A central mechanism: interaction

A term *interacts* with its context.
Some interactions are *well-behaved* (terminate), some do not.

Orthogonality

Consider a term t.
Set $t^{\perp}$ the set of contexts t interacts well with.
Set $t^{\perp\perp}$ the set of terms the contexts in $t^{\perp}$ interact well with.

The term in $t^{\perp\perp}$ behaves as t

Orthogonality is a classification principle.

An emerging logic

Suppose tu is of type A, π of type $A^\perp$.
$\rightsquigarrow$ tu and π interact well.

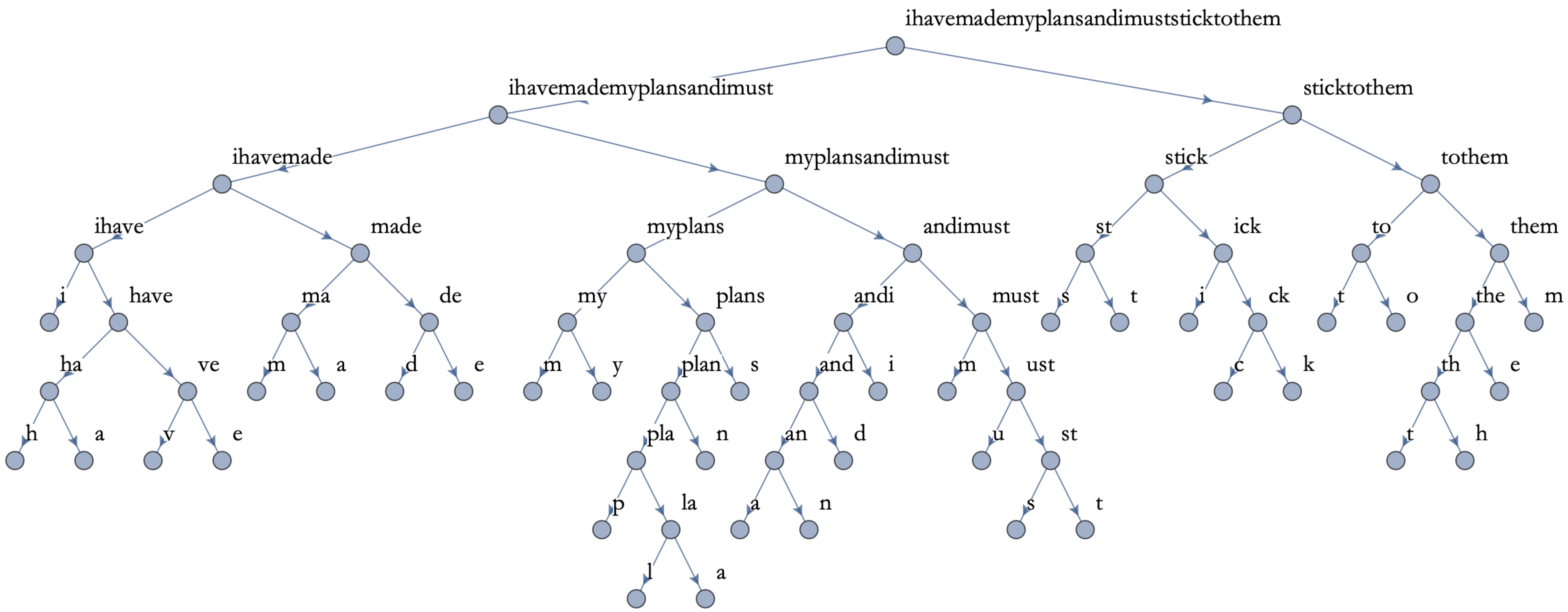
$$tu \mid \pi \quad \longrightarrow \quad t \mid u \bullet \pi$$

Hence, t interacts well with $u \bullet \pi$.
Suppose u is of type B, we can write $u \bullet \pi$ of type $B \bullet A^\perp$.
Hence t is of type $(B \bullet A^\perp)^\perp$, which we write $B \to A$.

A logic of implication, negation and conjunction emerges.

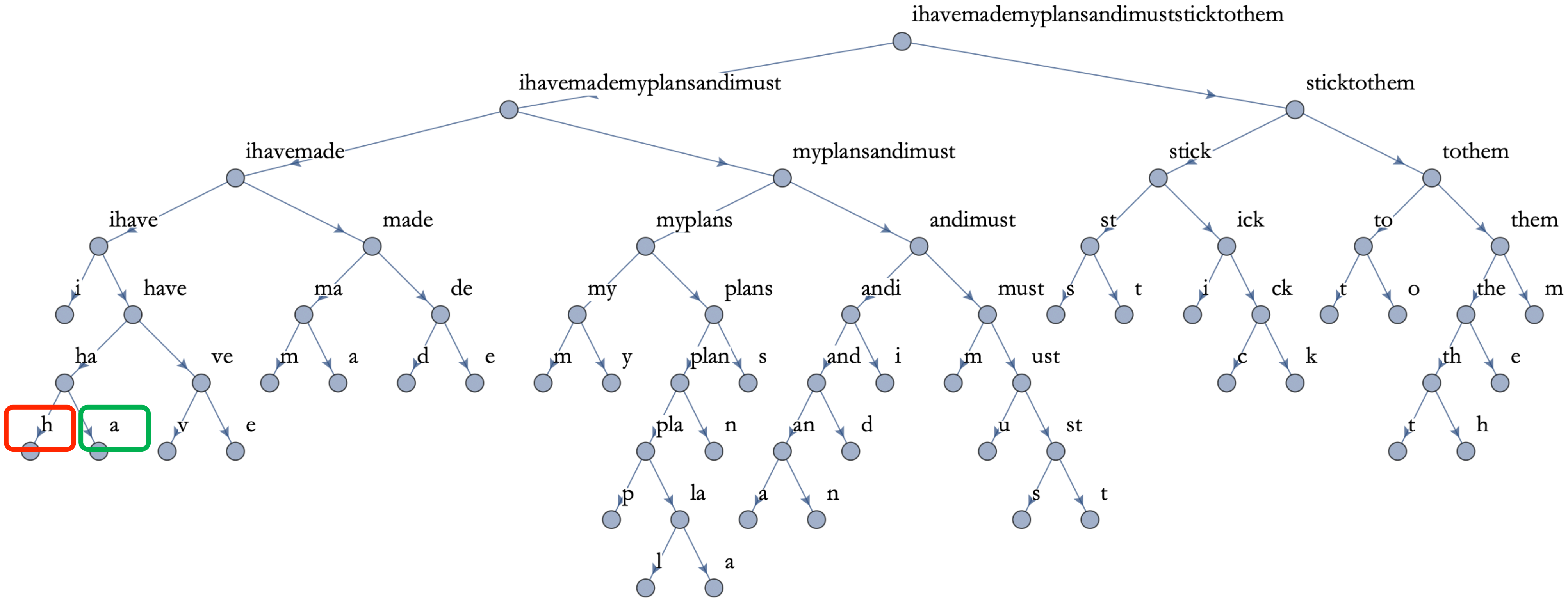
Take aways

1. *Interaction* as a notion of computation
   Between subterms of the term, not between agents
2. *Orthogonality* emerges from interaction
   $\rightsquigarrow$ a classification principle
3. Non-classical *Logic* emerging from orthogonality
   Connectives and their rules depend on the interaction

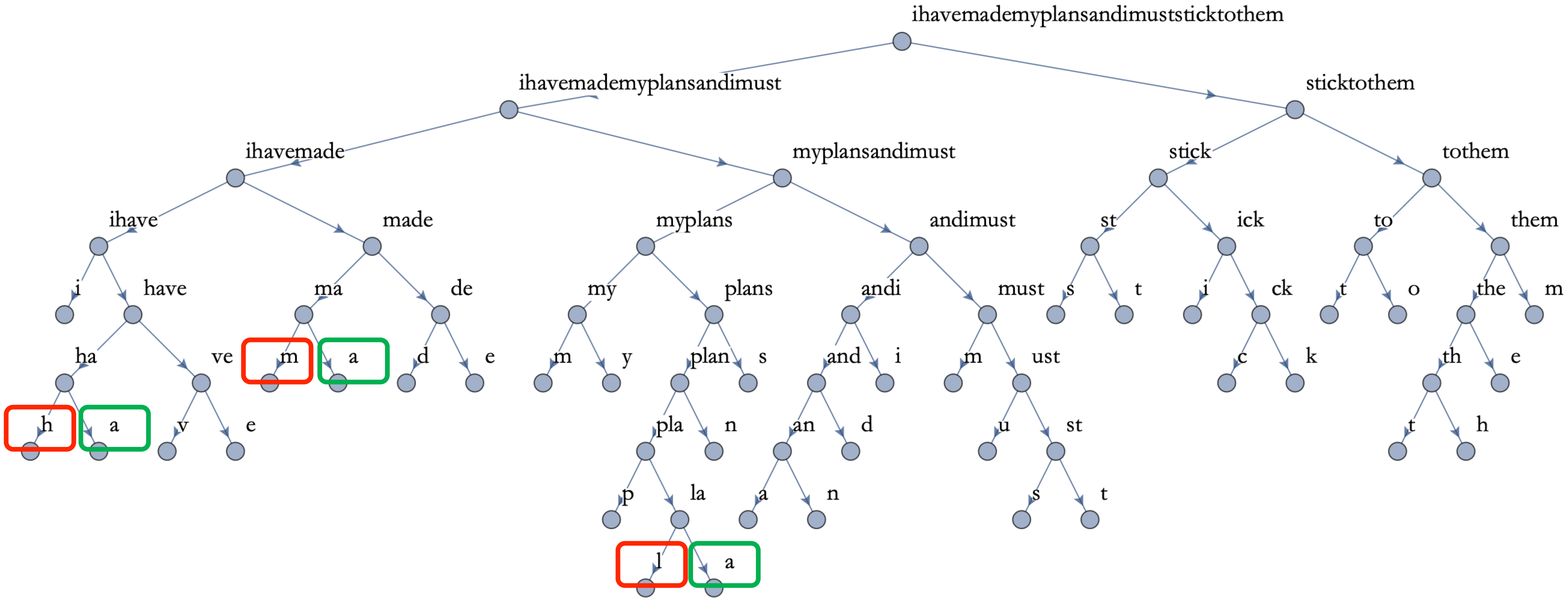


$$— \{h\} \to \perp = \{a\}$$

$$— \perp \leftarrow (\{h\} \to \perp) = \{h,m,l\}$$



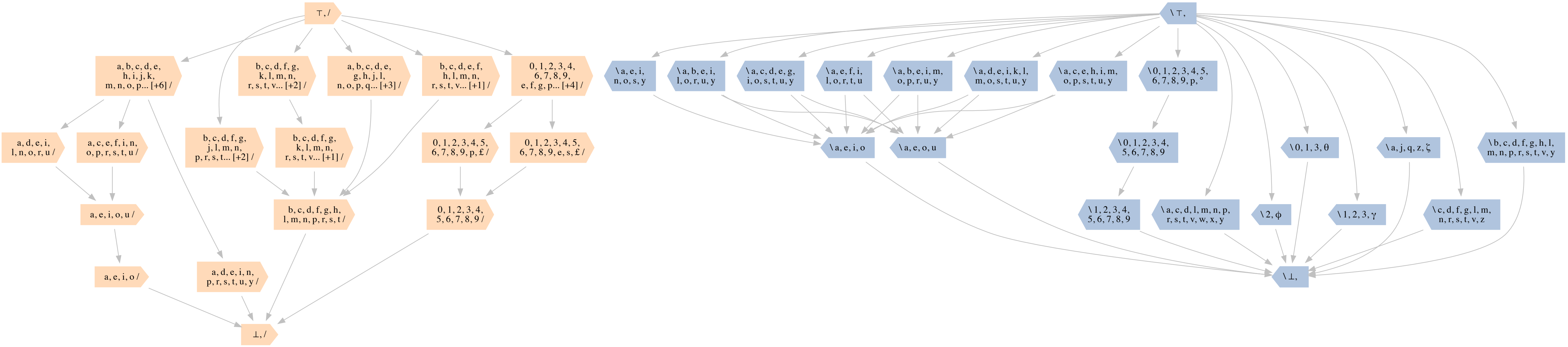
$$\text{---}\ \boxed{\{h\}} \to \perp = \boxed{\{a\}} \qquad\qquad \text{---}\ \perp \leftarrow (\{h\} \to \perp) = \{h,m,l\}$$

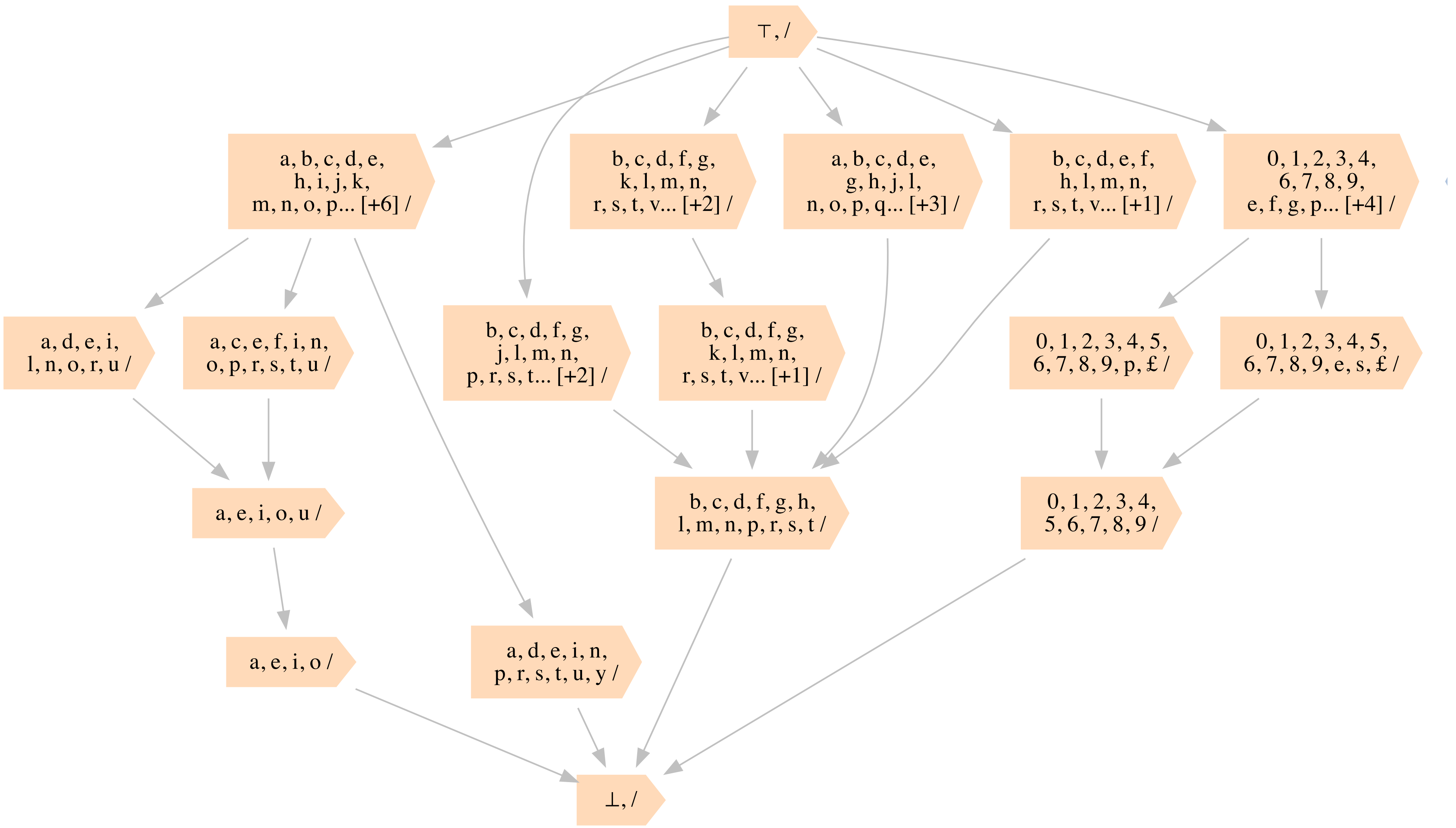


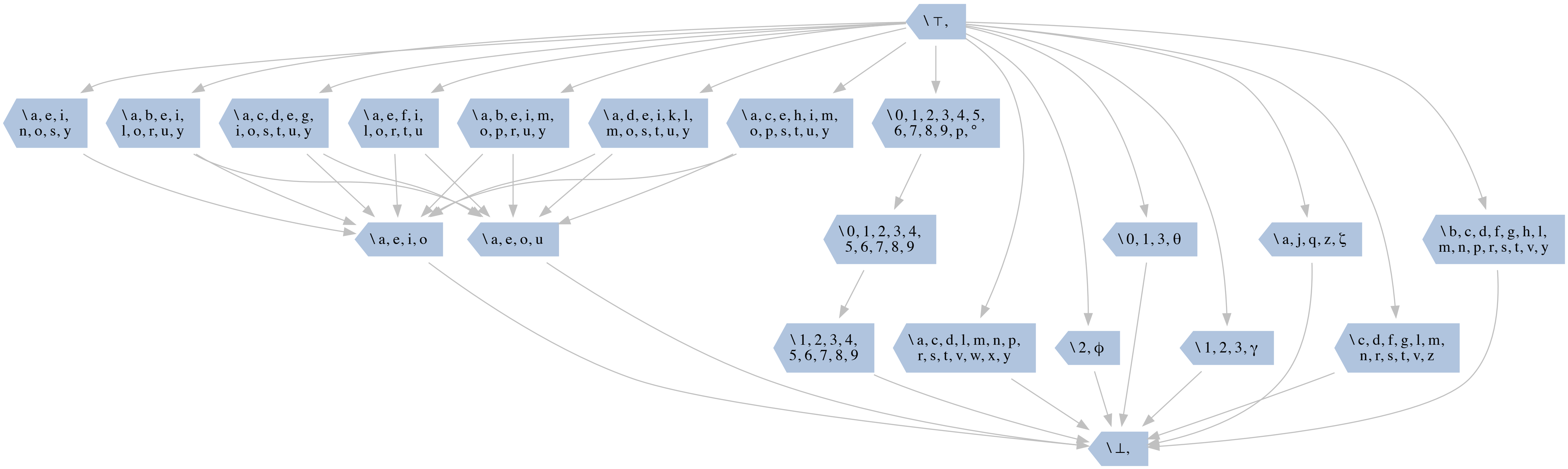
— $\{h\} \to \bot = \{a\}$

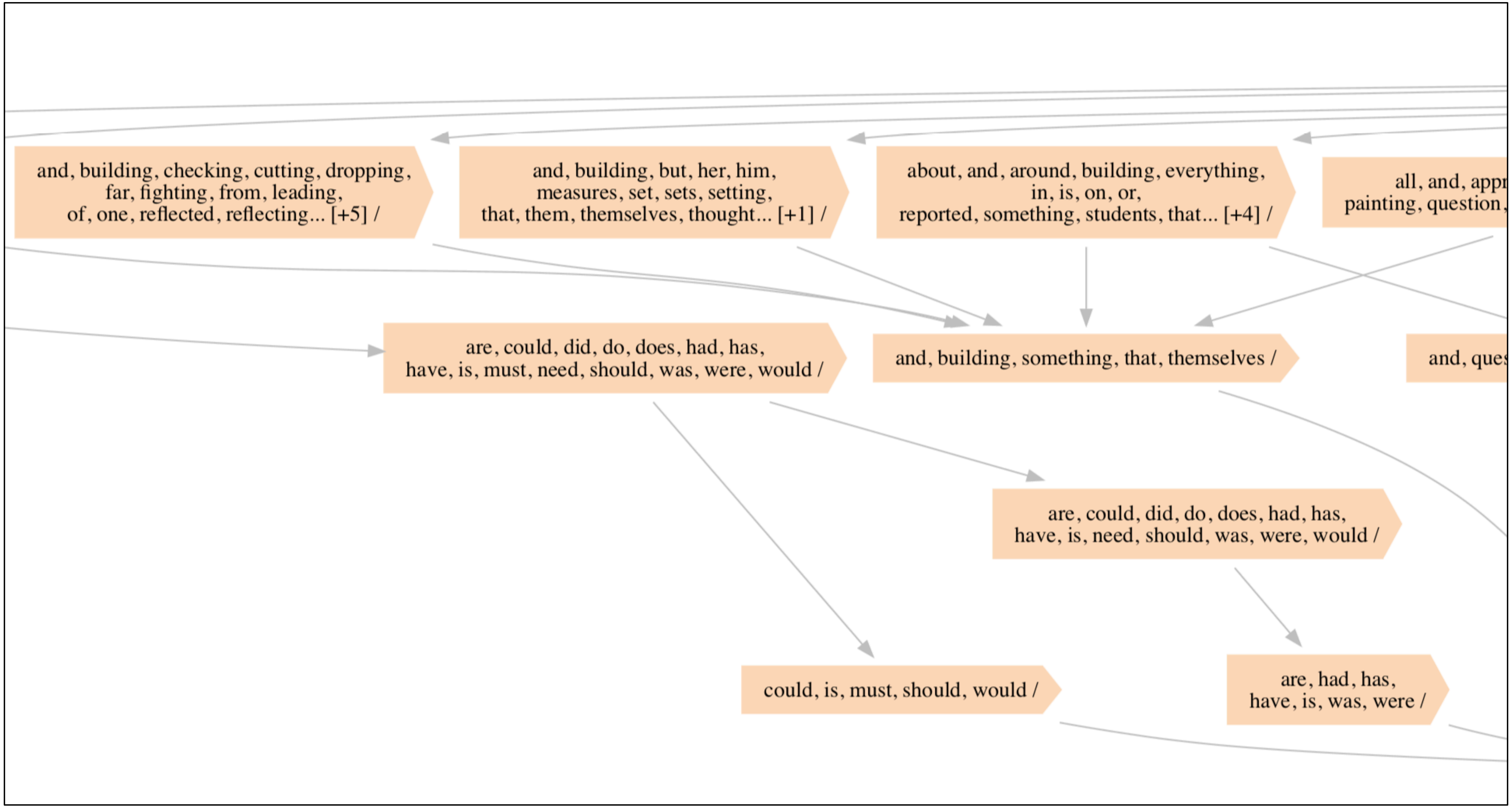
— $\bot \leftarrow \boxed{(\{h\} \to \bot)} = \boxed{\{h,m,l\}}$

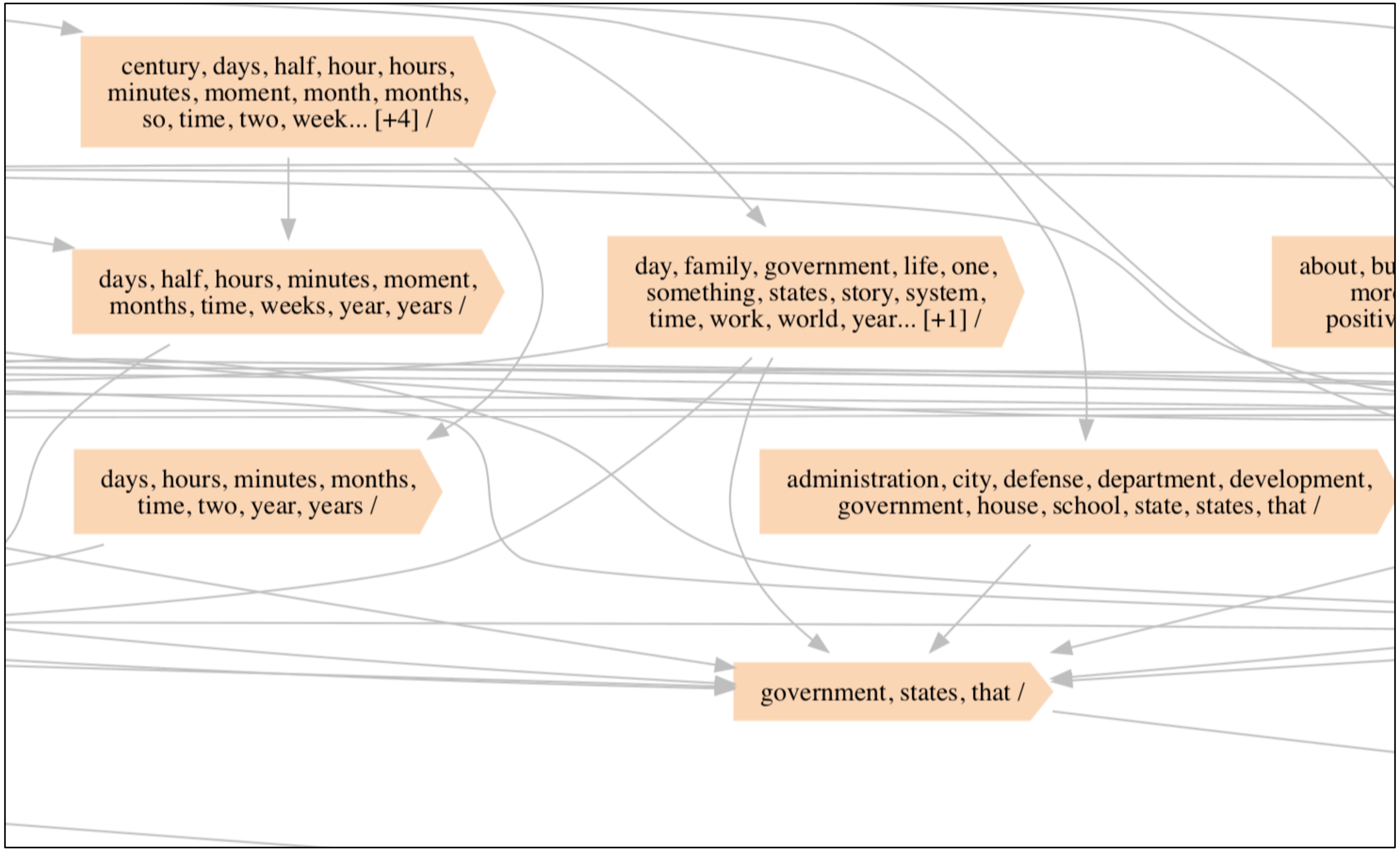
BOL	Term	BOR
()	0	()
()	1	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)
()	2	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)
()	3	()
()	4	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)
(1, 2, 3, 4, 5, 6, 7, 8, 9)	5	()
(1, 2, 3, 4, 5, 6, 7, 8, 9)	6	()
(1, 2, 3, 4, 5, 6, 7, 8, 9)	7	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)
(1, 2, 3, 4, 5, 6, 7, 8, 9)	8	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)
(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)	9	(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)
(a,)	a	(a,)
()	b	(b,)
(t,)	c	()
(d,)	d	(d,)
(a, e, i, o)	e	(e,)
(t,)	f	()
()	g	()
(h,)	h	()
(a, e, i, o)	i	(a, e, i)
()	j	(b, c, d, f, g, h, l, m, n, p, r, s, t)
(t,)	k	()


⊤, /
a, b, c, d, e, h, i, j, k, m, n, o, p... [+6] /
b, c, d, f, g, k, l, m, n, r, s, t, v... [+2] /
a, b, c, d, e, g, h, j, l, n, o, p, q... [+3] /
b, c, d, e, f, h, l, m, n, r, s, t, v... [+1] /
0, 1, 2, 3, 4, 6, 7, 8, 9, e, f, g, p... [+4] /
a, d, e, i, l, n, o, r, u /
a, c, e, f, i, n, o, p, r, s, t, u /
b, c, d, f, g, j, l, m, n, p, r, s, t... [+2] /
b, c, d, f, g, k, l, m, n, r, s, t, v... [+1] /
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, p, £ /
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, e, s, £ /
a, e, i, o, u /
b, c, d, f, g, h, l, m, n, p, r, s, t /
0, 1, 2, 3, 4, 5, 6, 7, 8, 9 /
a, e, i, o /
a, d, e, i, n, p, r, s, t, u, y /
⊥, /
\ ⊤,
\ a, e, i, n, o, s, y
\ a, b, e, i, l, o, r, u, y
\ a, c, d, e, g, i, o, s, t, u, y
\ a, e, f, i, l, o, r, t, u
\ a, b, e, i, m, o, p, r, u, y
\ a, d, e, i, k, l, m, o, s, t, u, y
\ a, c, e, h, i, m, o, p, s, t, u, y
\ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, p, °
\ a, e, i, o
\ a, e, o, u
\ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
\ 0, 1, 3, θ
\ a, j, q, z, ζ
\ b, c, d, f, g, h, l, m, n, p, r, s, t, v, y
\ 1, 2, 3, 4, 5, 6, 7, 8, 9
\ a, c, d, l, m, n, p, r, s, t, v, w, x, y
\ 2, ϕ
\ 1, 2, 3, γ
\ c, d, f, g, l, m, n, r, s, t, v, z
\ ⊥,

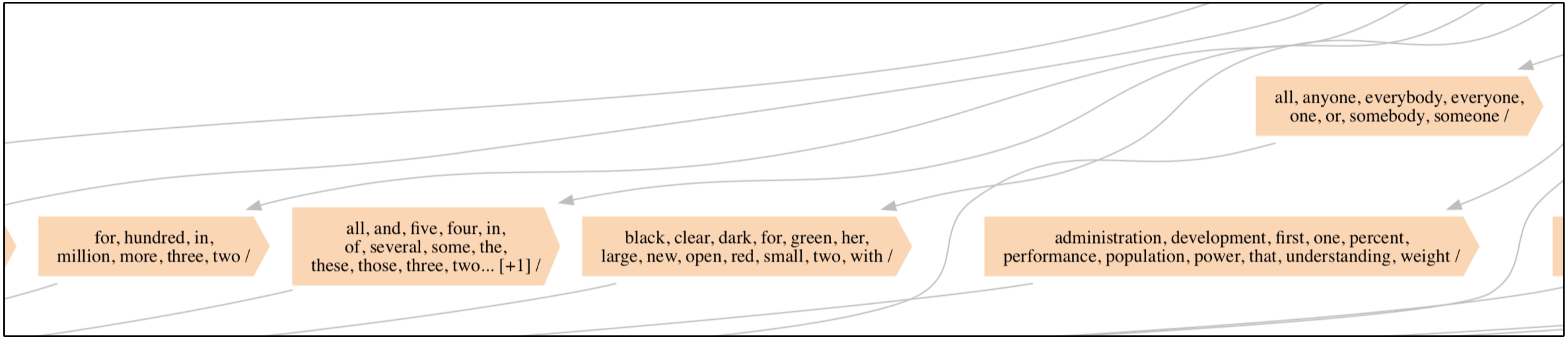

⊤, /
a, b, c, d, e, h, i, j, k, m, n, o, p... [+6] /
b, c, d, f, g, k, l, m, n, r, s, t, v... [+2] /
a, b, c, d, e, g, h, j, l, n, o, p, q... [+3] /
b, c, d, e, f, h, l, m, n, r, s, t, v... [+1] /
0, 1, 2, 3, 4, 6, 7, 8, 9, e, f, g, p... [+4] /
a, d, e, i, l, n, o, r, u /
a, c, e, f, i, n, o, p, r, s, t, u /
b, c, d, f, g, j, l, m, n, p, r, s, t... [+2] /
b, c, d, f, g, k, l, m, n, r, s, t, v... [+1] /
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, p, £ /
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, e, s, £ /
a, e, i, o, u /
b, c, d, f, g, h, l, m, n, p, r, s, t /
0, 1, 2, 3, 4, 5, 6, 7, 8, 9 /
a, e, i, o /
a, d, e, i, n, p, r, s, t, u, y /
⊥, /

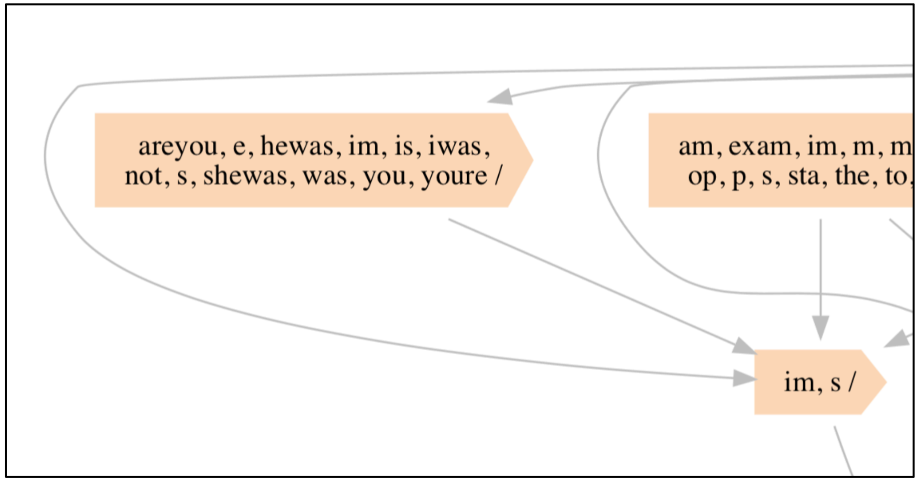

\ ⊤,
\ a, e, i, n, o, s, y
\ a, b, e, i, l, o, r, u, y
\ a, c, d, e, g, i, o, s, t, u, y
\ a, e, f, i, l, o, r, t, u
\ a, b, e, i, m, o, p, r, u, y
\ a, d, e, i, k, l, m, o, s, t, u, y
\ a, c, e, h, i, m, o, p, s, t, u, y
\ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, p, °
\ a, e, i, o
\ a, e, o, u
\ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
\ 0, 1, 3, θ
\ a, j, q, z, ζ
\ b, c, d, f, g, h, l, m, n, p, r, s, t, v, y
\ 1, 2, 3, 4, 5, 6, 7, 8, 9
\ a, c, d, l, m, n, p, r, s, t, v, w, x, y
\ 2, ϕ
\ 1, 2, 3, γ
\ c, d, f, g, l, m, n, r, s, t, v, z
\ ⊥,

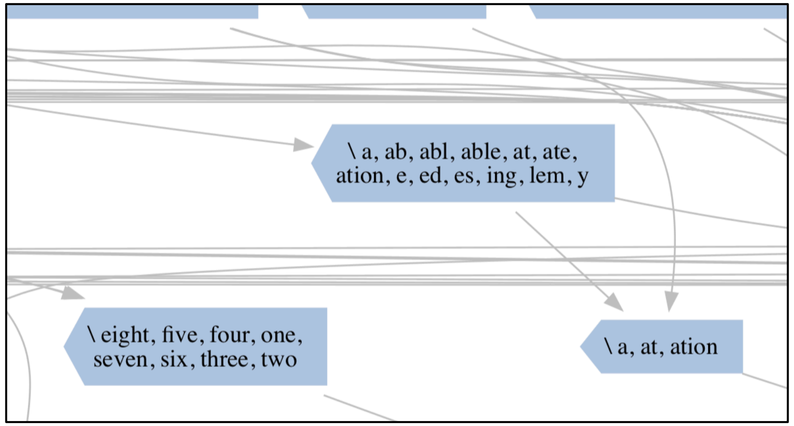

and, building, checking, cutting, dropping,
far, fighting, from, leading,
of, one, reflected, reflecting... [+5] /
and, building, but, her, him,
measures, set, sets, setting,
that, them, themselves, thought... [+1] /
about, and, around, building, everything,
in, is, on, or,
reported, something, students, that... [+4] /
are, could, did, do, does, had, has,
have, is, must, need, should, was, were, would /
and, building, something, that, themselves /
are, could, did, do, does, had, has,
have, is, need, should, was, were, would /
could, is, must, should, would /
are, had, has,
have, is, was, were /

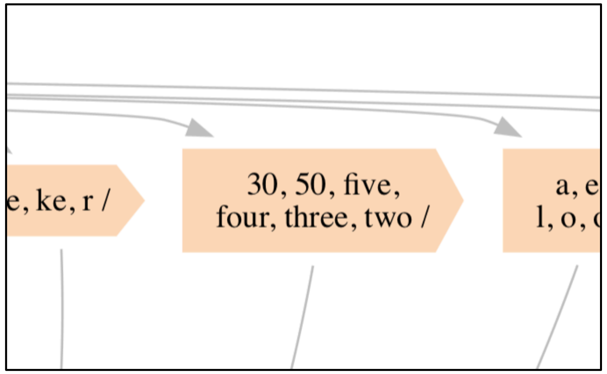

century, days, half, hour, hours, minutes, moment, month, months, so, time, two, week... [+4] /
days, half, hours, minutes, moment, months, time, weeks, year, years /
day, family, government, life, one, something, states, story, system, time, work, world, year... [+1] /
days, hours, minutes, months, time, two, year, years /
administration, city, defense, department, development, government, house, school, state, states, that /
government, states, that /

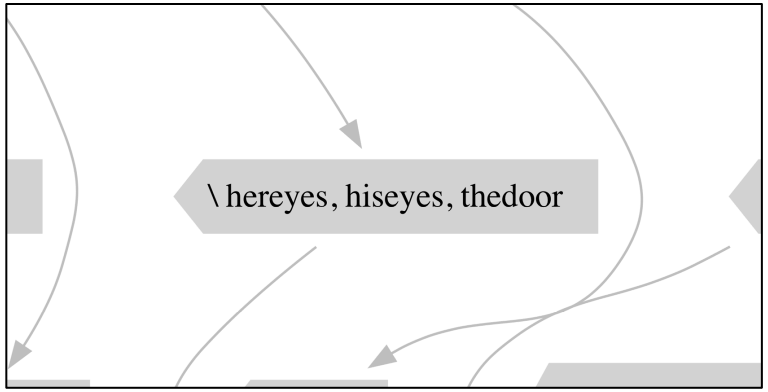
PROOF OF CONCEPT




areyou, e, hewas, im, is, iwas,
not, s, shewas, was, you, youre /
am, exam, im, m, m
op, p, s, sta, the, to,
im, s /


\ a, ab, abl, able, at, ate,
ation, e, ed, es, ing, lem, y
\ eight, five, four, one,
seven, six, three, two
\ a, at, ation


e, ke, r /
30, 50, five,
four, three, two /
a, e
l, o,


\ hereyes, hiseyes, thedoor

- Cultural (collective) instead of cognitive (individual) framework
- Alternative way of connecting (theoretical) computer science to meaning
- Possible links between DNN models and theoretical computer science
- Neither connectionist, not symbolic, but still logical
- Not limited to natural language

5th International Conference on the
History and Philosophy of Computing (HaPoC)
28-30 Oct 2019 Bergamo (Italy)

The Logic of Language

An alternative logical approach to machine learning based on the case of Natural Language Processing

Juan Luis Gastaldi
HPM – ETH Zürich
juan.luis.gastaldi@gess.ethz.ch

Luc Pellissier
IRIF – Paris Diderot
pellissier@irif.fr

This project has received funding from the
European Union's Horizon 2020 research and innovation programme
under grant agreement No 839730