Deep Grammar

From Natural Language Processing to Artificial Intelligence

Jonathan Mugan, PhD

@jmugan

Data Day Texas

January 14, 2017

It is not my aim to surprise or shock you – but the simplest way I can summarize is to say that there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until – in a visible future – the range of problems they can handle will be coextensive with the range to which the human mind has been applied.

“ It is not my aim to surprise or shock you – but the simplest way I can summarize is to say that there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until – in a visible future – the range of problems they can handle will be coextensive with the range to which the human mind has been applied. ”

Herbert Simon, 1957

" It is not my aim to surprise or shock you – but the simplest way I can summarize is to say that there are now in the world machines that think, that learn and t
these things is going to inc
the range of problems the
range to which the human

Herbert Simon, 1957

So what is going on here?

There is no actual understanding.

AI has gotten smarter
- especially with deep learning, but
- computers can't read or converse intelligently

This is disappointing because we want to

- Interact with our world using natural language
  - Current chatbots are an embarrassment
- Have computers read all those documents out there
  - So they can retrieve the best ones, answer our questions, and summarize what is new

Deep Grammar

To understand language, computers need to understand the world

They need to be able to answer questions like:

Why can you pull a wagon with a string but not push it? -- Minsky

Why is it unusual that a gymnast competed with one leg? -- Schank

Why does it only rain outside?

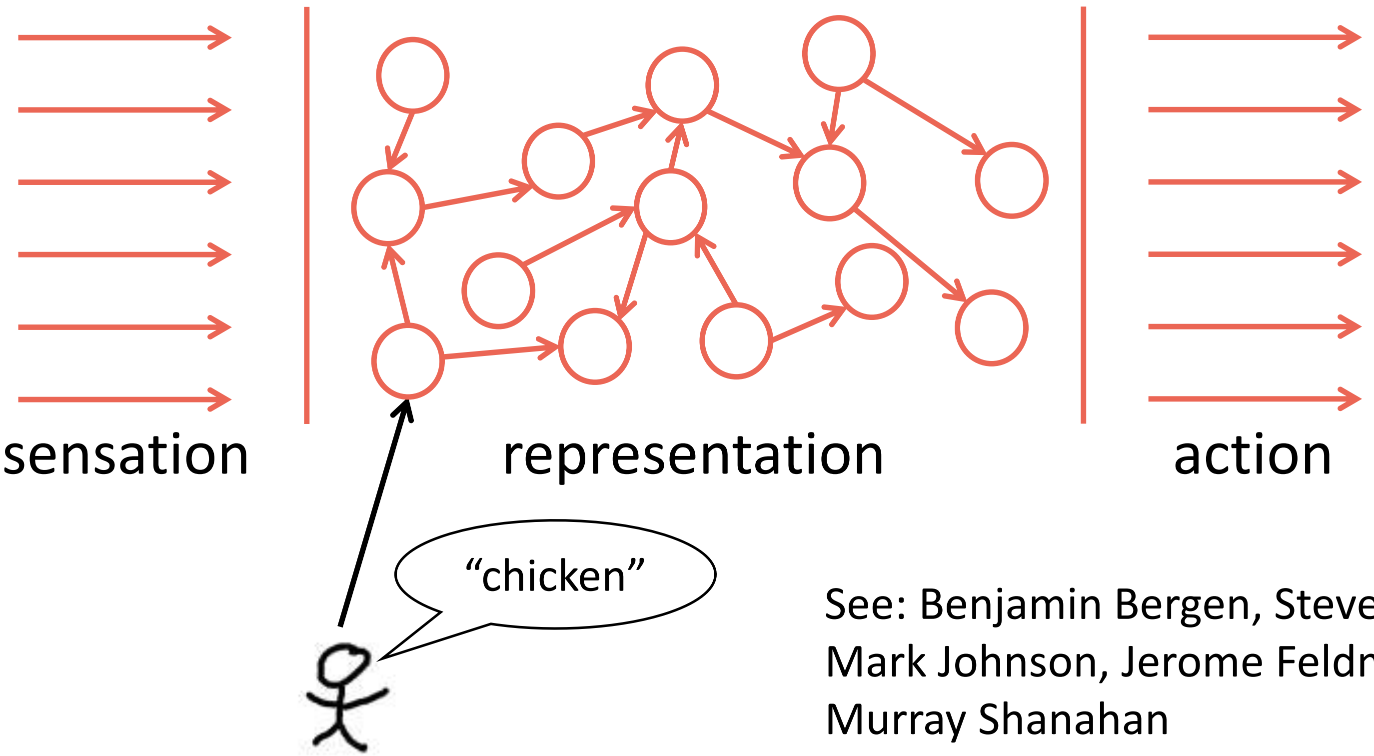
If a book is on a table, and you push the table, what happens?

If Bob went to the junkyard, is he at the airport?

Deep Grammar

Grounded Understanding

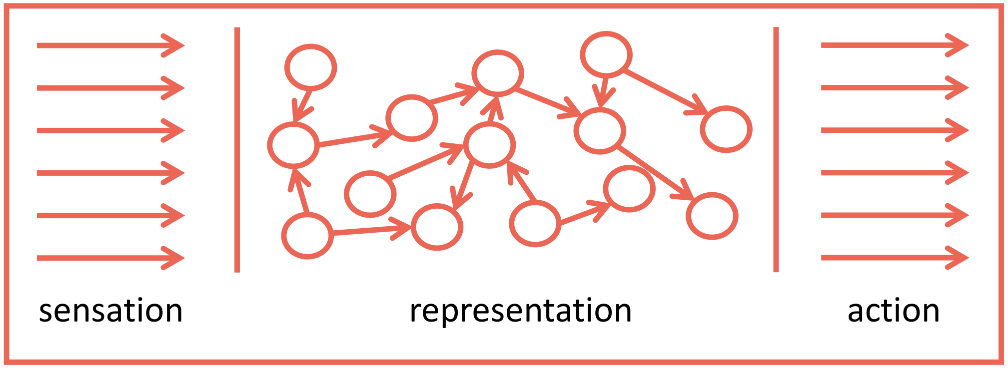
- We understand language in a way that is grounded in sensation and action.



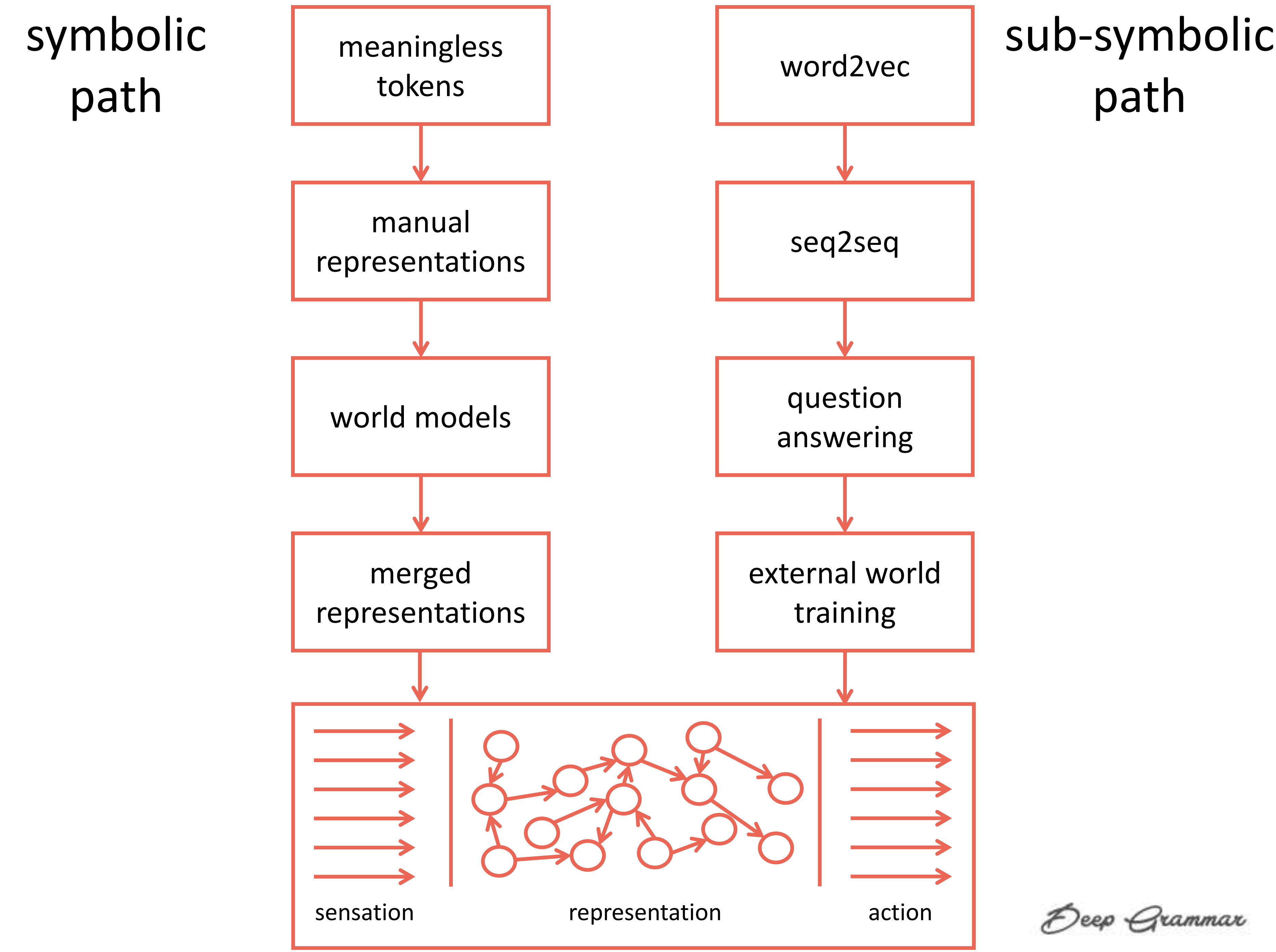
See: Benjamin Bergen, Steven Pinker, Mark Johnson, Jerome Feldman, and Murray Shanahan

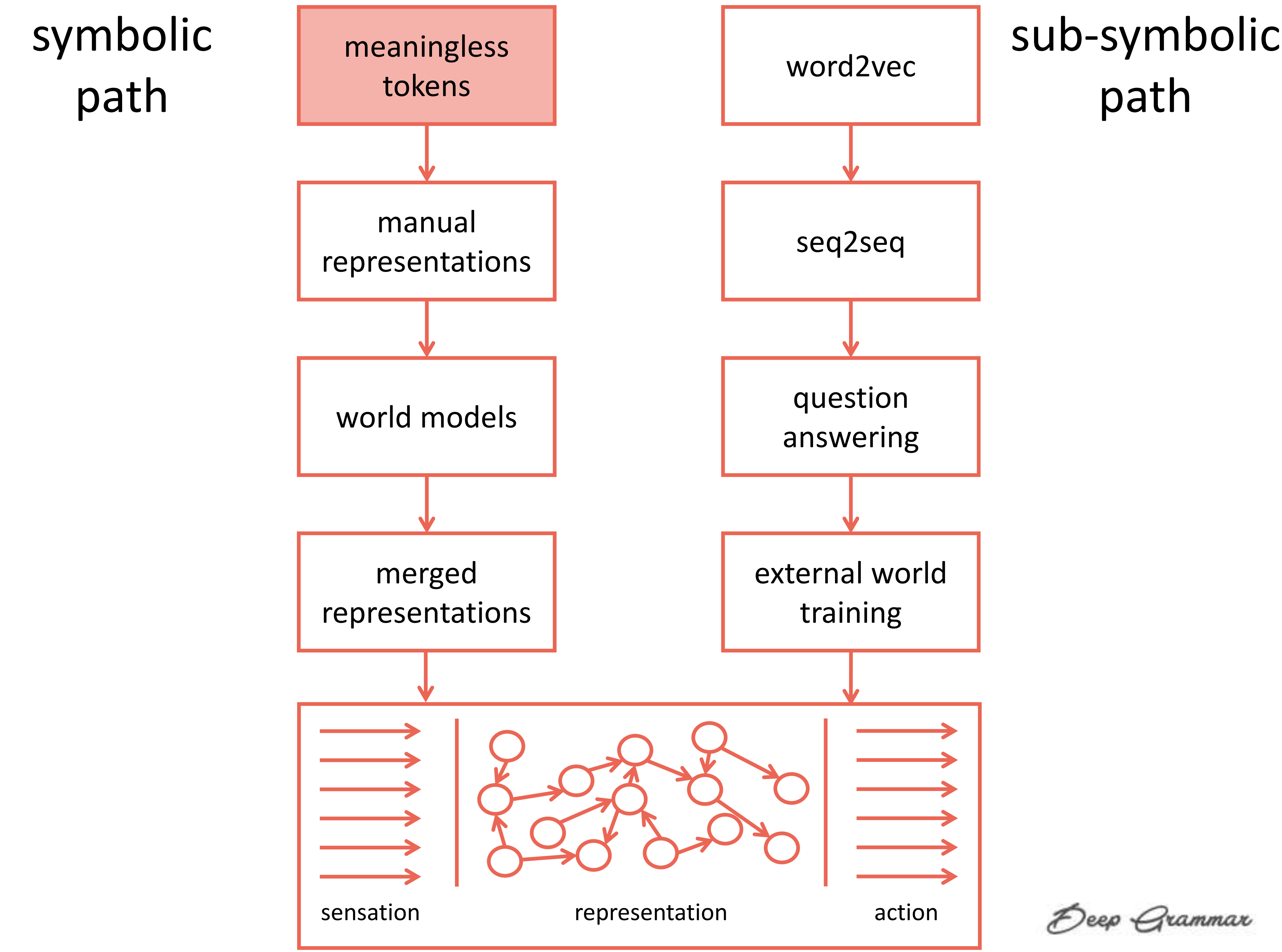
- When someone says “chicken,” we map that to our experience with chickens.
- We understand each other because we have had similar experiences.
- This is the kind of understanding that computers need.

Two paths to meaning:



Deep Grammar


symbolic path
meaningless tokens
manual representations
world models
merged representations
sub-symbolic path
word2vec
seq2seq
question answering
external world training
sensation
representation
action
Deep Grammar


symbolic path
meaningless tokens
manual representations
world models
merged representations
word2vec
seq2seq
question answering
external world training
sub-symbolic path
sensation
representation
action
Deep Grammar

Bag-of-words representation

Treat words as arbitrary symbols and look at their frequencies.
“dog bit man” will be the same as “man bit dog”

Consider a vocabulary of 50,000 words where:

- “aardvark” is position 0
- “ate” is position 2
- “zoo” is position 49,999

A bag-of-words can be a vector with 50,000 dimensions.

“The aardvark ate the zoo.” = [1,0,1, ..., 0, 1]

We can do a little better and count how often the words occur.
tf: term frequency, how often does the word occur.

“The aardvark ate the aardvark by the zoo.” = [2,0,1, ..., 0, 1]

Give rare words a boost

We can get fancier and say that rare words are more important than common words for characterizing documents.

Multiply each entry by a measure of how common it is in the corpus.

idf: inverse document frequency

idf(term, document) = log(num. documents / num. with term)

10 documents, only 1 has “aardvark” and 5 have “zoo” and 5 have “ate”

tf-idf: tf * idf

“The aardvark ate the aardvark by the zoo.” =[4.6,0,0.7, ..., 0, 0.7]

Called a vector space model. You can throw these vectors into any classifier, or find similar documents based on similar vectors.

Topic Modeling (LDA)

Latent Dirichlet Allocation

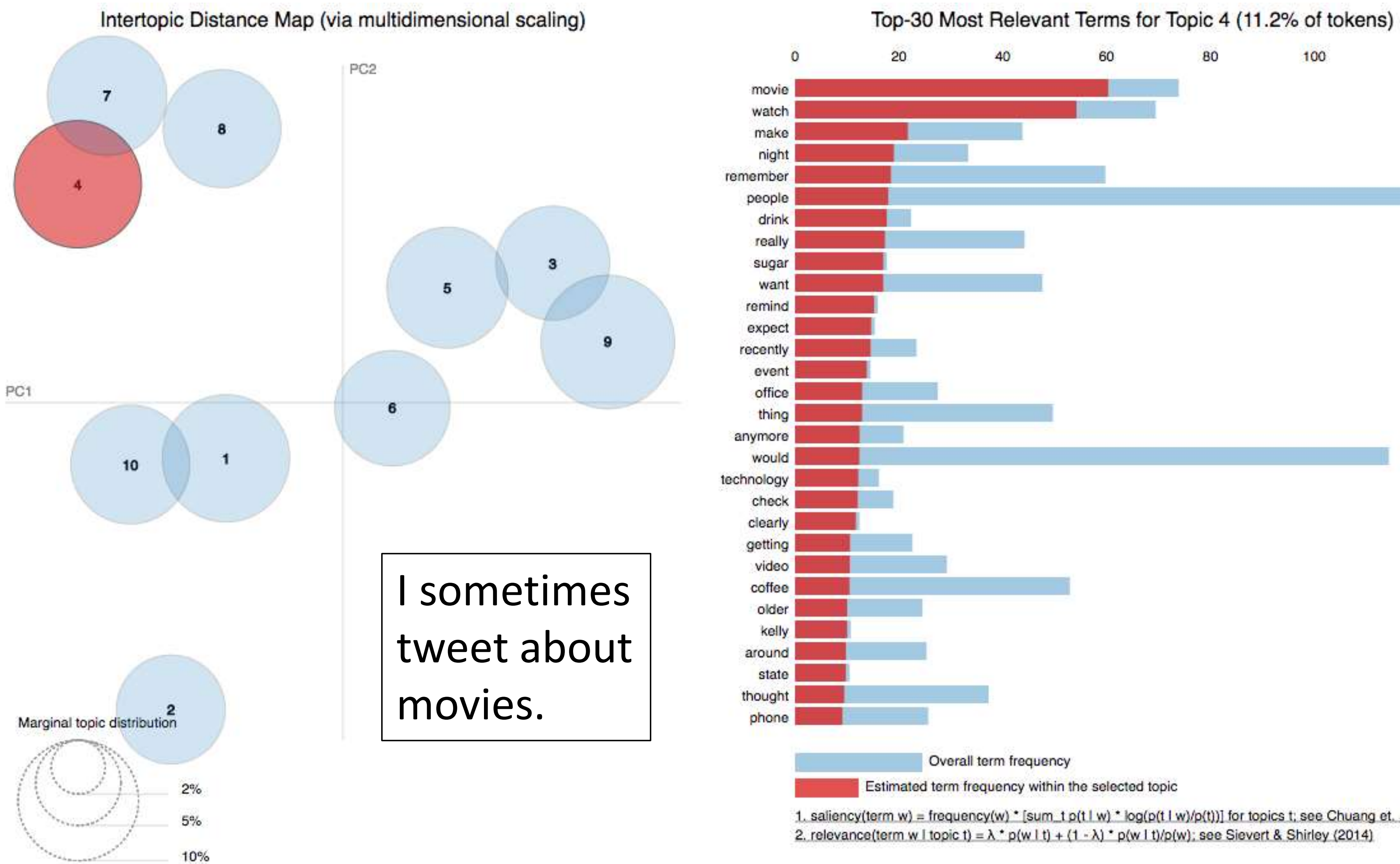
- You pick the number of topics
- Each topic is a distribution over words
- Each document is a distribution over topics

Easy to do in gensim

https://radimrehurek.com/gensim/models/ldamodel.html

LDA of my tweets shown in pyLDAvis

https://github.com/bmabey/pyLDAvis



I sometimes tweet about movies.

Sentiment analysis: How the author feels about the text

We can do sentiment analysis using labeled data and meaningless tokens, with supervised learning over tf-idf.

We can also do sentiment analysis by adding the first hint of meaning: some words are positive and some words are negative.

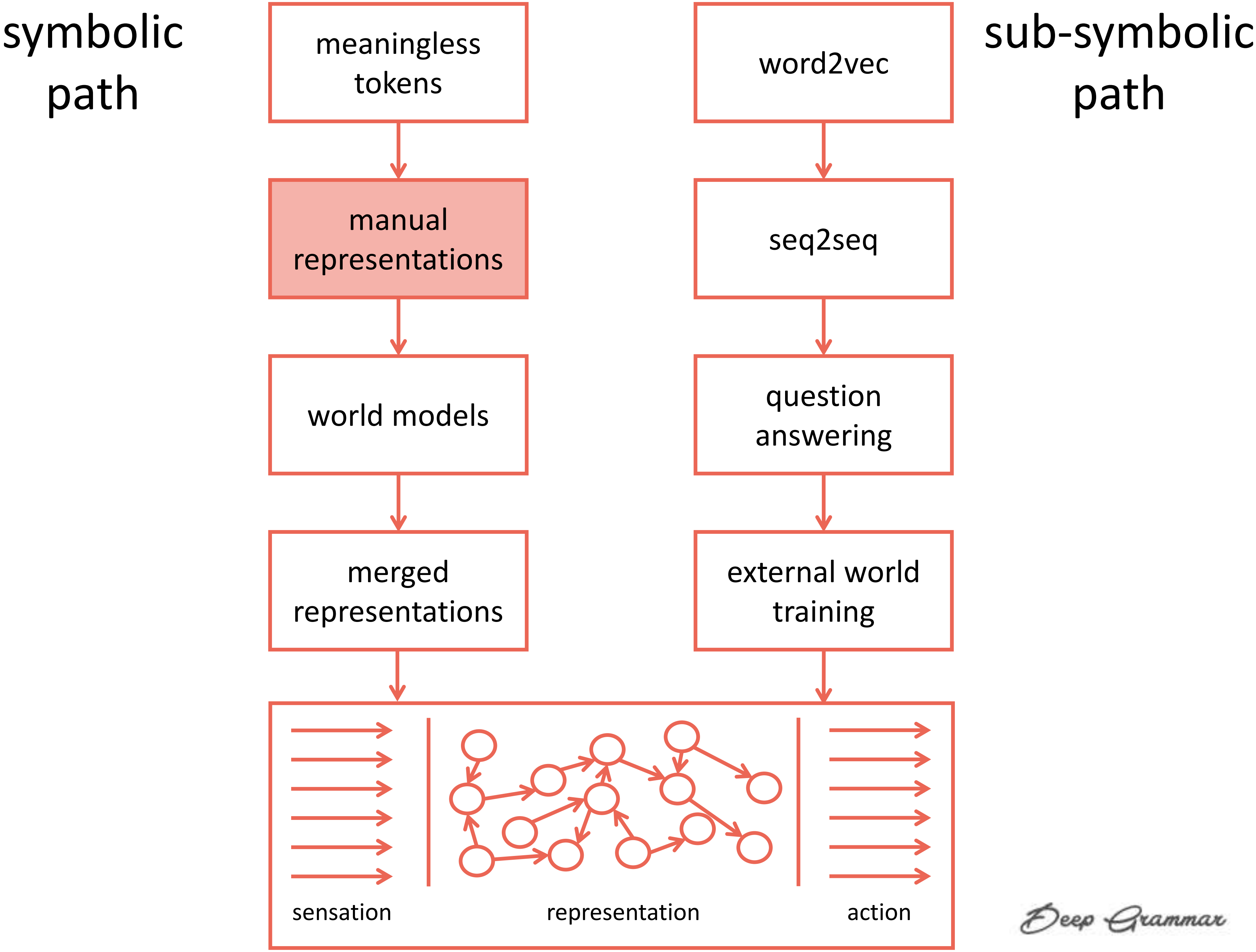
Sentiment dictionary: word list with sentiment values associated with all words that might indicate sentiment

```
...
happy: +2
...
joyful: +2
...
pain: -3
painful: -3
...
```

"I went to the junkyard and was happy to see joyful people."

One such word list is VADER https://github.com/cjhutto/vaderSentiment

Deep Grammar


symbolic path
meaningless tokens
manual representations
world models
merged representations
word2vec
seq2seq
question answering
external world training
sub-symbolic path
sensation
representation
action
Deep Grammar

Manually Constructing Representations

We tell the computer what things mean by manually specifying relationships between symbols

1. Stores meaning using predefined relationships
2. Maps multiple ways of writing something to the same representation

Allows us to code what the machine should do for a relatively small number of representations

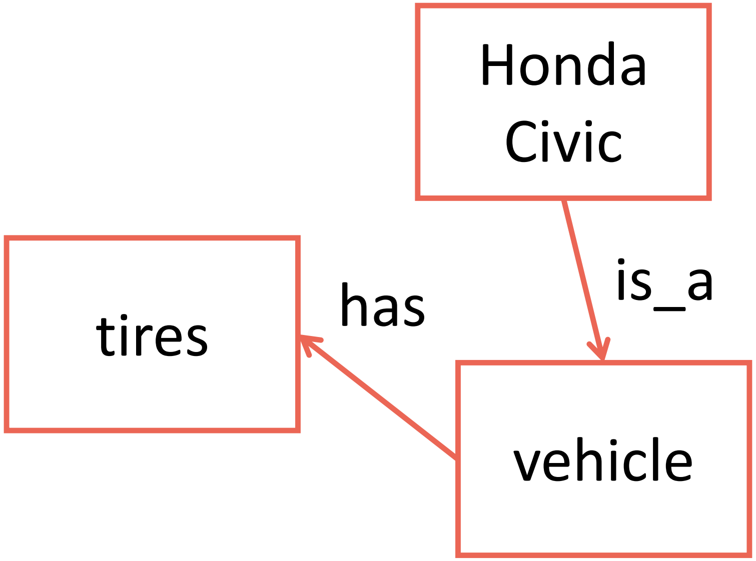
Manually Constructing Representations

We tell the computer what things mean by manually specifying relationships between symbols

1. Stores meaning using predefined relationships
2. Maps multiple ways of writing something to the same representation

Allows us to code what the machine should do for a relatively small number of representations

Who might be in the market for tires?



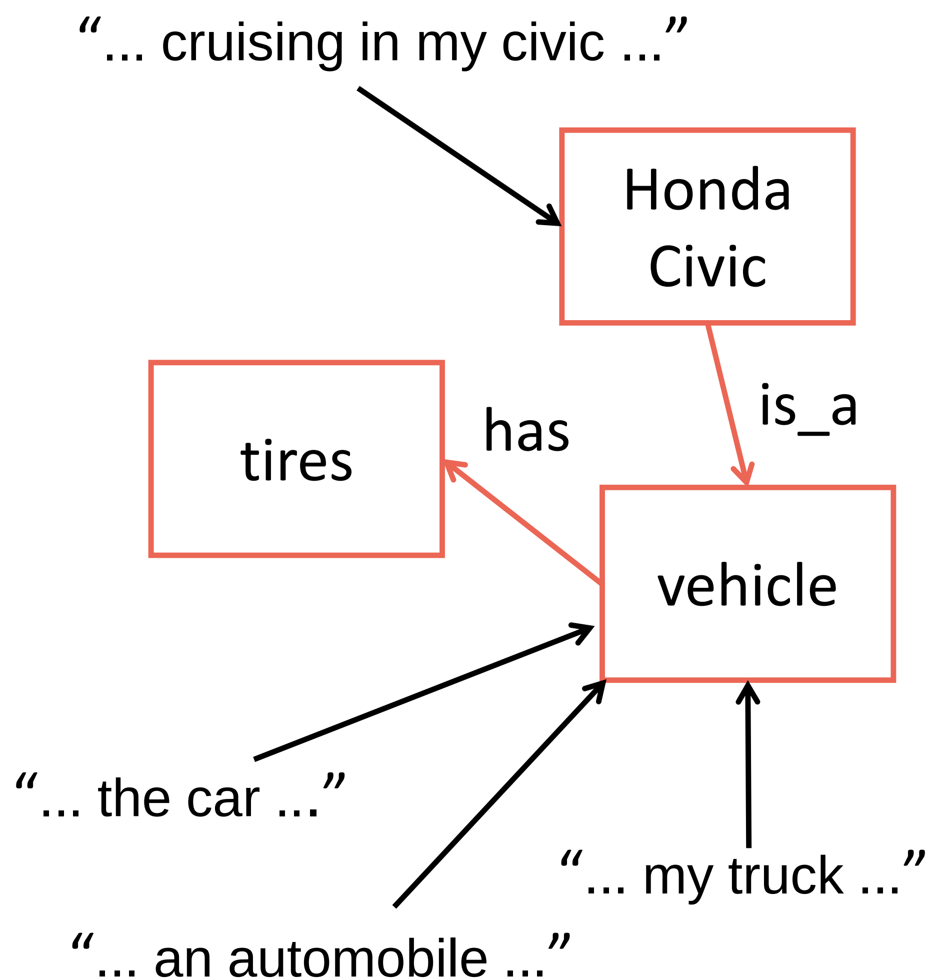
Manually Constructing Representations

We tell the computer what things mean by manually specifying relationships between symbols

1. Stores meaning using predefined relationships
2. Maps multiple ways of writing something to the same representation

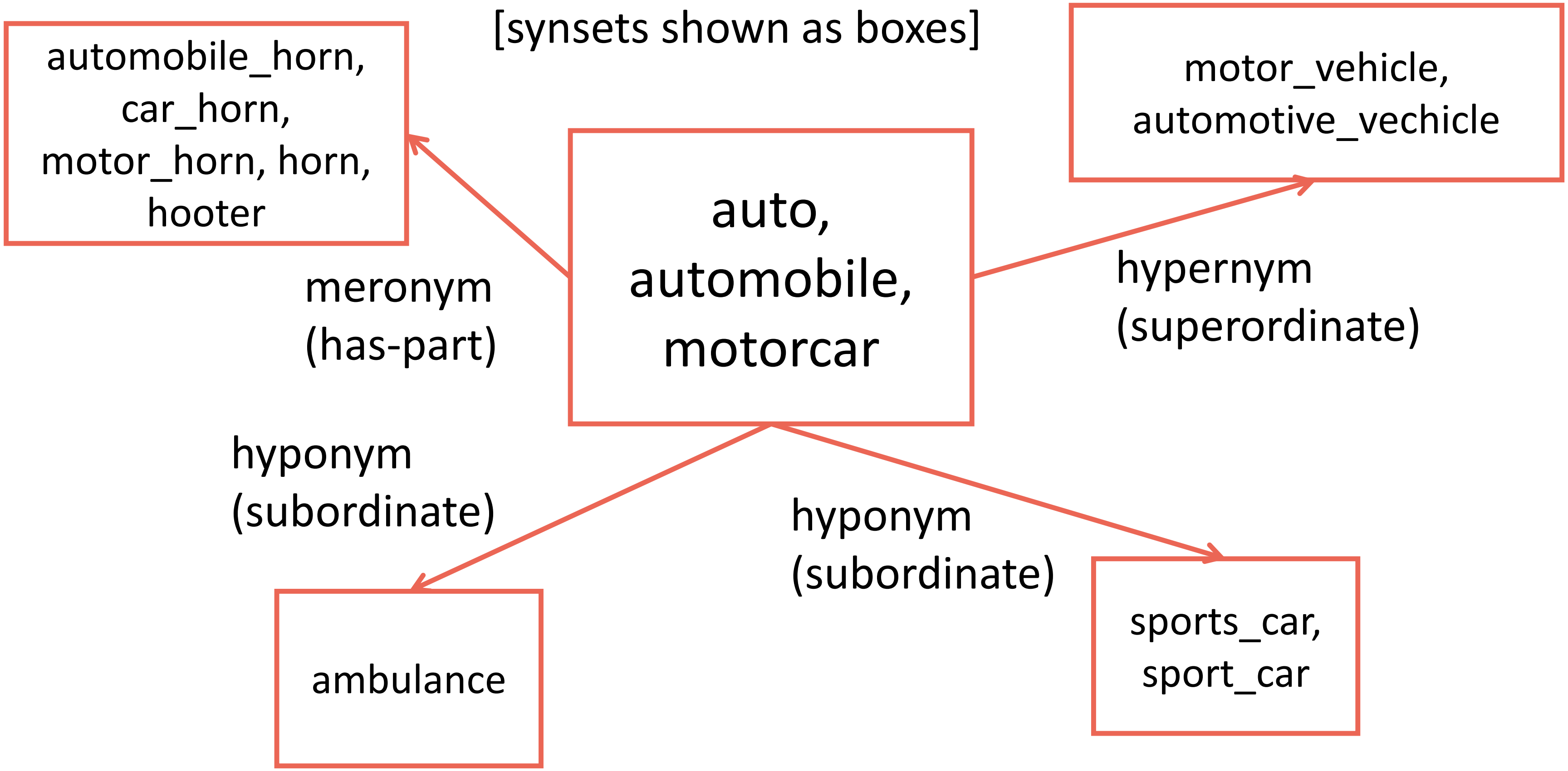
Allows us to code what the machine should do for a relatively small number of representations

Who might be in the market for tires?



WordNet

- Organizes sets of words into *synsets*, which are meanings
- A word can be a member of more than one synset, such as “bank”



WordNet Search http://wordnetweb.princeton.edu/perl/webwn

WordNet

- Organizes sets of words into *synsets*, which are meanings
- A word can be a member of more than one synset, such as “bank”

Other relations:

- has-instance
    - car : Honda Civic
- has-member
    - team : stringer (like “first stringer”)
- antonym
- ... a few more

WordNet Search http://wordnetweb.princeton.edu/perl/webwn

FrameNet

- More integrative than WordNet: represents situations
- One example is a child's birthday party, another is Commerce_buy
  - situations have slots (roles) that are filled
- Frames are triggered by keywords in text (more or less)

Commerce_buy triggered by the words: buy, buyer, client, purchase, or purchaser

Getting

Roles: Buyer, Goods, Money, Place (where bought), ...

Commerce_buy

Inherits from

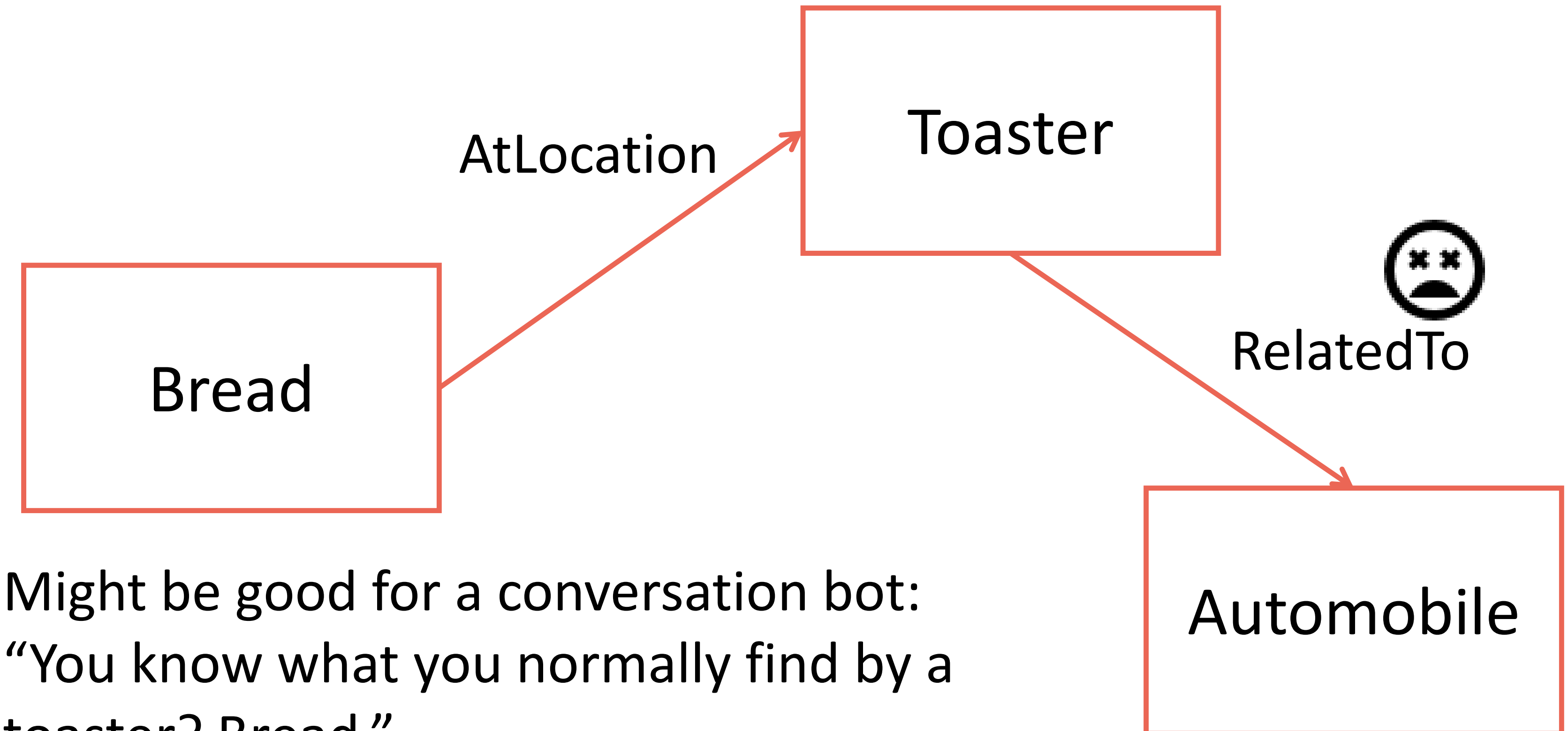
Commerce_buy indicates a change of possession, but we need a world model to actually change a state.

FrameNet: https://framenet.icsi.berkeley.edu/fndrupal/IntroPage

Commerce_buy: https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Commerce_buy

ConceptNet

- Provides commonsense linkages between words
- My experience: too shallow and haphazard to be useful



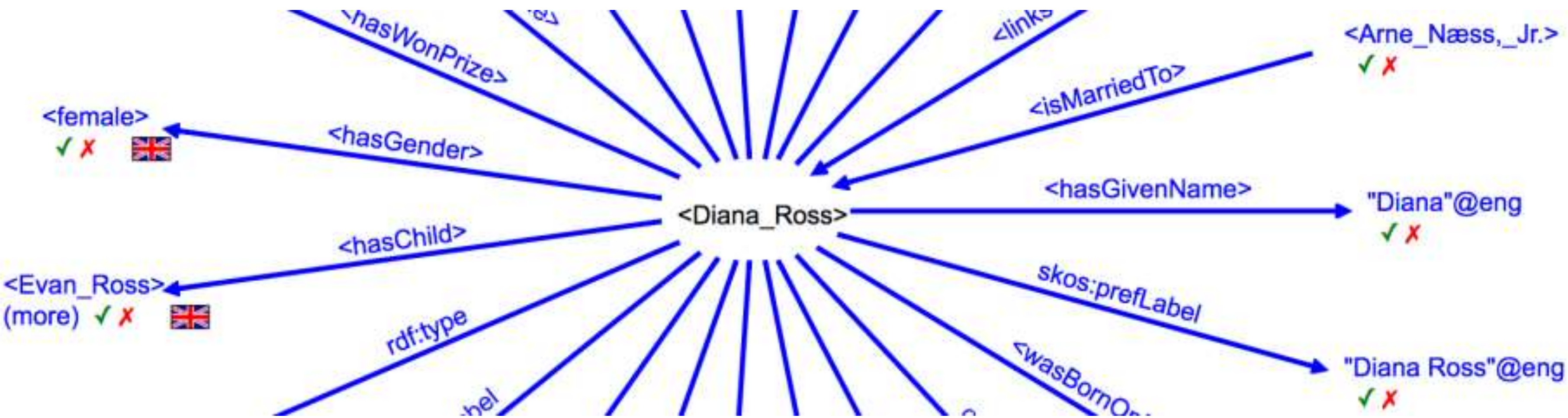
Might be good for a conversation bot:
“You know what you normally find by a toaster? Bread.”

ConceptNet: http://conceptnet.io/

Deep Grammar

YAGO: Yet Another Great Ontology

- Built on WordNet and DBpedia
  - http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/
  - DBpedia has a machine readable page for each Wikipedia page
- Used by IBM Watson to play Jeopardy!
  - Big on named entities, like entertainers
- Browse
  - https://gate.d5.mpi-inf.mpg.de/webyago3spotlx/Browser



SUMO: Suggested Upper Merged Ontology

Deep: organizes concepts down to the lowest level
http://www.adampease.org/OP/

Example: *cooking* is a type of *making* that is a type of *intentional process* that is a type of *process* that is a *physical thing* that is an *entity*.

There is also YAGO-SUMO that merges the low-level organization of SUMO with the instance information of YAGO.http://people.mpi-inf.mpg.de/~gdemelo/yagosumo/

Deep Grammar

Image Schemas

Image schemas are representations of human experience that are common across cultures [Feldman, 2006]

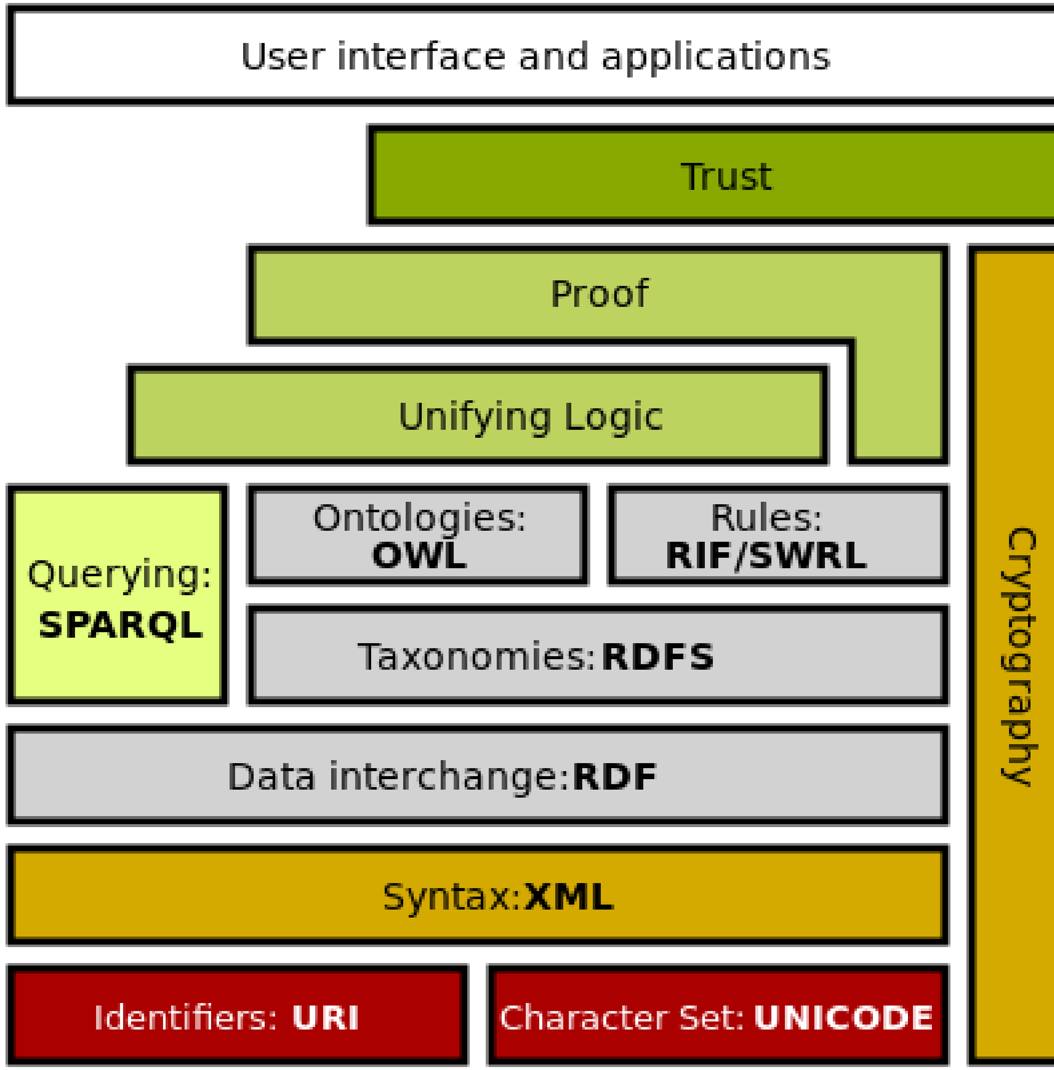
Humans use image schemas comprehend spatial arrangements and movements in space [Mandler, 2004]

Examples of image schemas include *path*, *containment*, *blockage*, and *attraction* [Johnson, 1987]

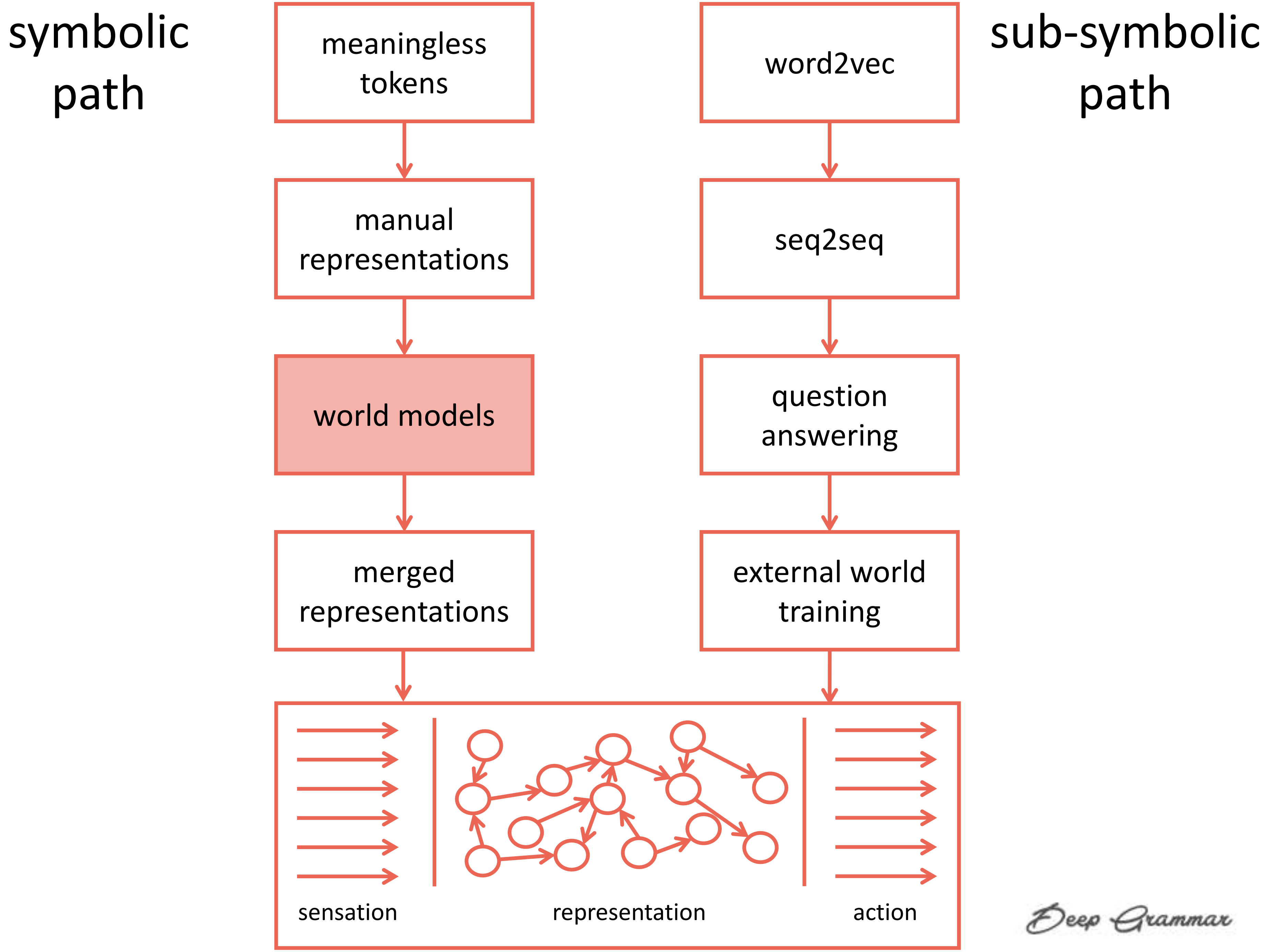
Abstract concepts such as romantic relationships and arguments are represented as metaphors to this kind of experience [Lakoff and Johnson, 1980]

Semantic Web (Linked Data)

- Broad, but not organized or deep
- Way too complicated
- May eventually be streamlined (e.g. JSON-LD), and it could be very cool if it gets linked with deeper, better organized data
- Tools to map text:
  - FRED
  - DBpedia Spotlight
  - Pikes



Semantic Web Layer Cake Spaghetti Monster


symbolic path
meaningless tokens
manual representations
world models
merged representations
sub-symbolic path
word2vec
seq2seq
question answering
external world training
sensation
representation
action
Deep Grammar

World models

Computers need causal models of how the world works and how we interact with it.

- People don't say everything to get a message across, just what is not covered by our shared conception
- Most efficient way to encode our shared conception is a model

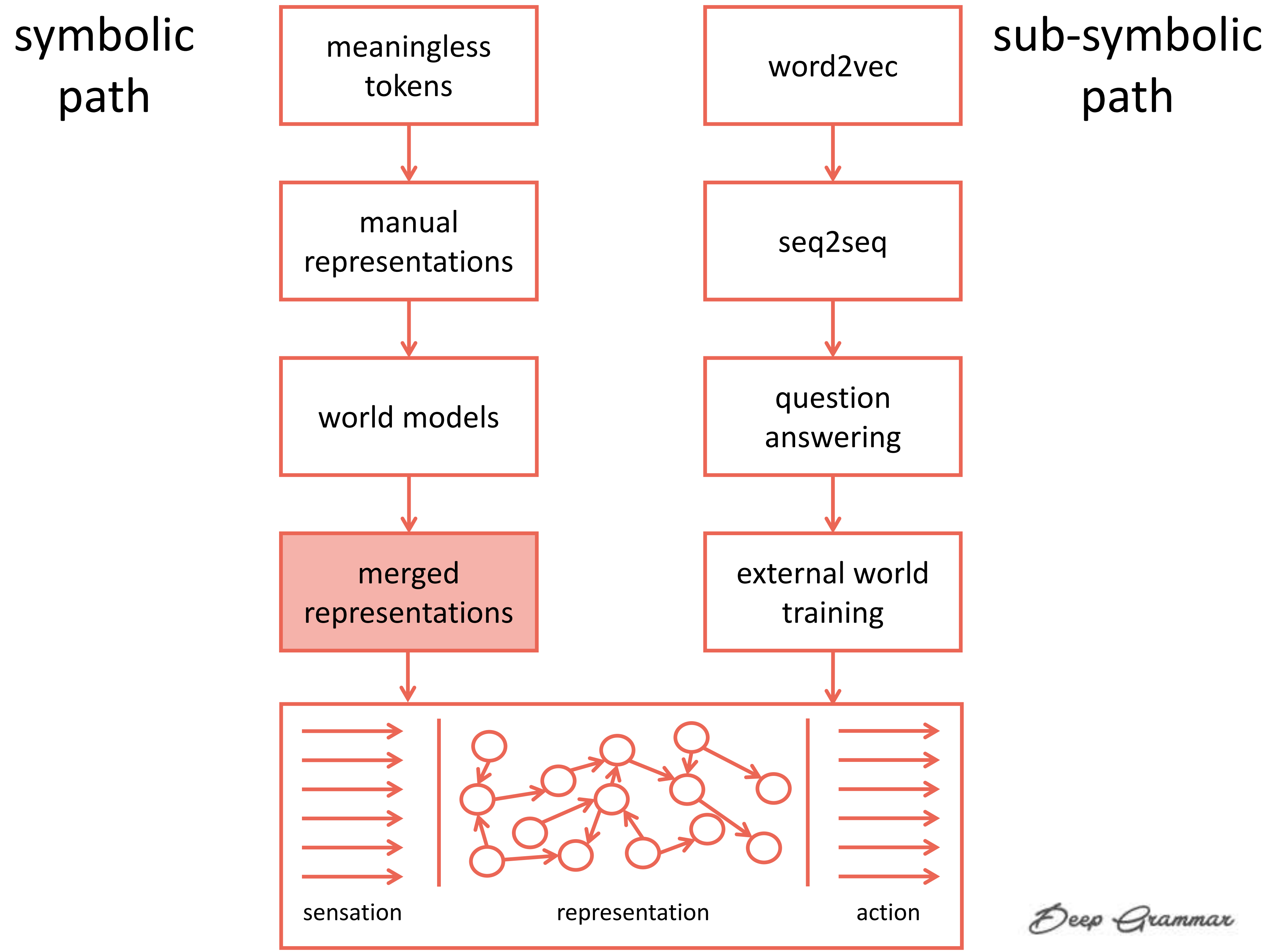
Models express how the world changes based on events

- Recall the Commerce_buy frame
- Afterward, one person has more money and another person has less
- Read such inferences right off the model

Deep Grammar

Dimensions of models

- Probabilistic
  - Deterministic compared with stochastic
  - E.g., logic compared with probabilistic programming
- Factor state
  - whole states compared with using variables
  - E.g., finite automata compared with dynamic Bayesian networks
- Relational
  - Propositional logic compared with first-order logic
  - E.g., Bayesian networks compared with Markov logic networks
- Concurrent
  - Model one thing compared with multiple things
  - E.g., finite automata compared with Petri Nets
- Temporal
  - Static compared with dynamic
  - E.g., Bayesian networks compared with dynamic Bayesian networks


symbolic
path
meaningless
tokens
manual
representations
world models
merged
representations
word2vec
seq2seq
question
answering
external world
training
sub-symbolic
path
sensation
representation
action
Deep Grammar

Merge representations with models

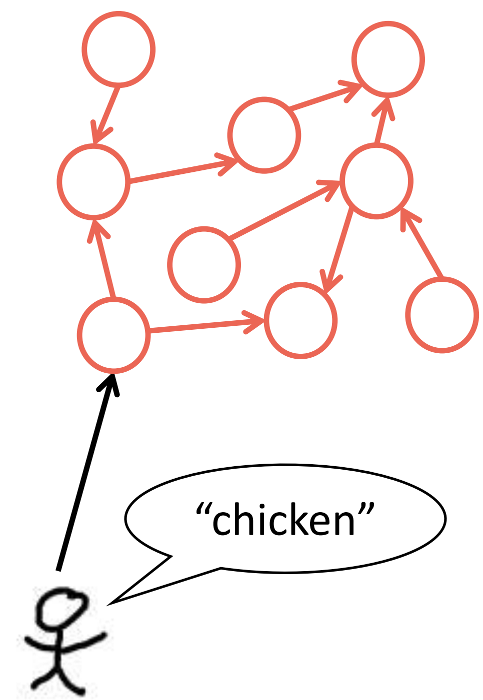
The final step on this path is to create a robust model around rich representations.

Why does it only rain outside?
A roof blocks the path of things from above.

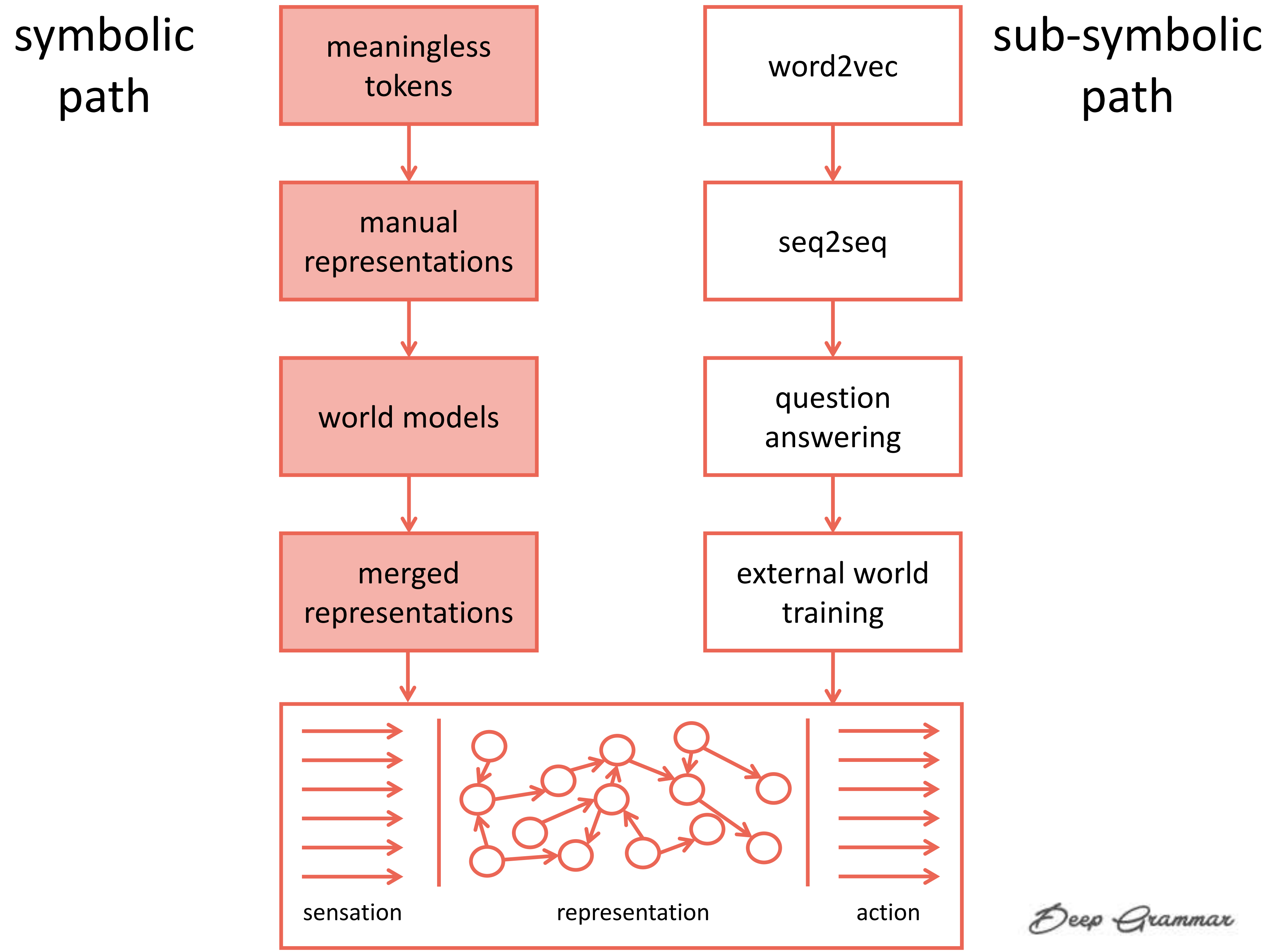
Explain your conception of conductivity?
Electrons are small spheres, and electricity is small spheres going through a tube. Conductivity is how little blockage is in the tube.

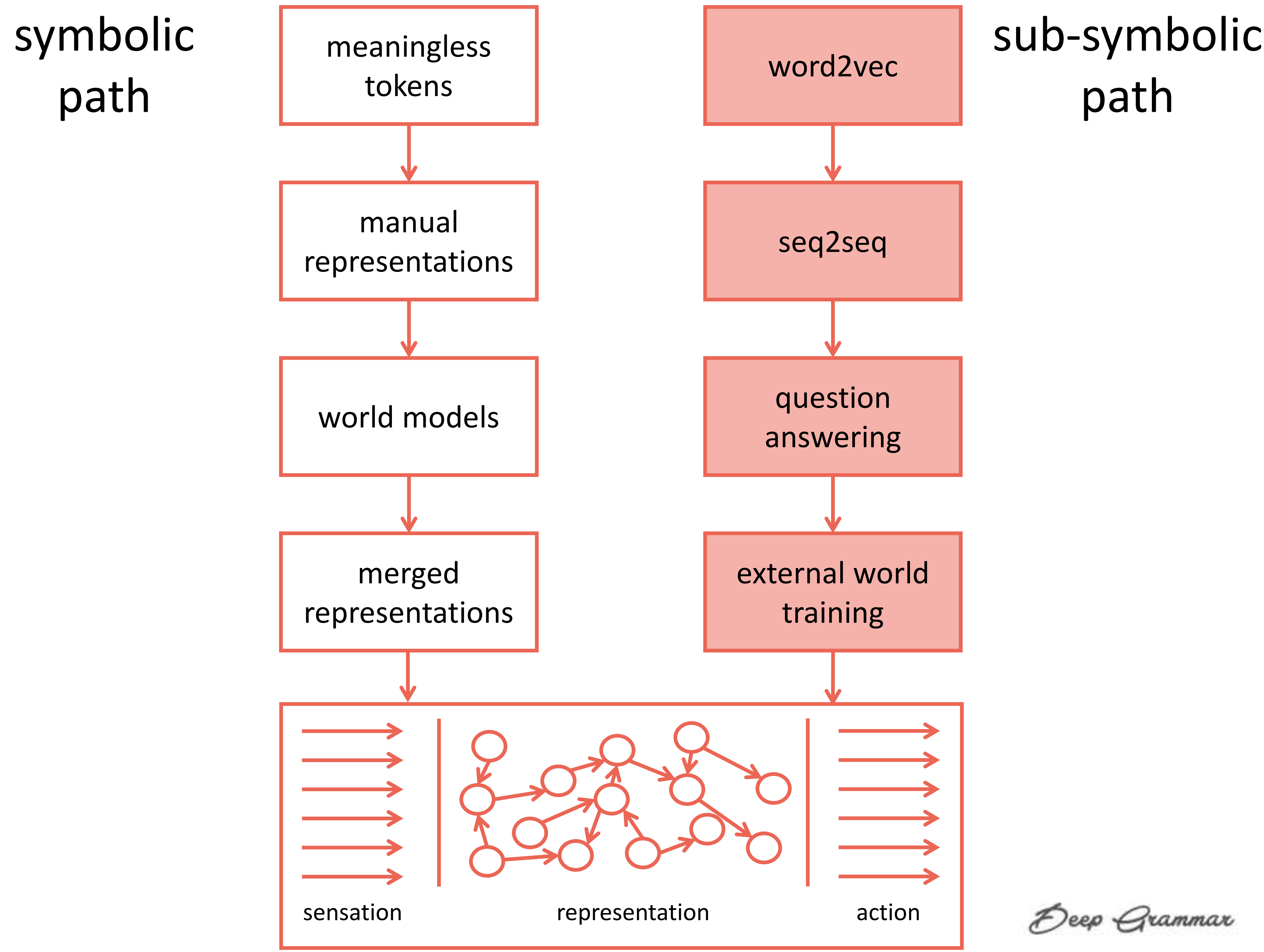
Cyc has a model that uses representations, but it is not clear if logic is sufficiently supple.

The meaning of the word "chicken" is everything explicitly stated in the representation and everything that can be inferred from the world model.



Deep Grammar


symbolic path
meaningless tokens
manual representations
world models
merged representations
word2vec
seq2seq
question answering
external world training
sub-symbolic path
sensation
representation
action
Deep Grammar


symbolic path
meaningless tokens
manual representations
world models
merged representations
word2vec
seq2seq
question answering
external world training
sub-symbolic path
sensation
representation
action
Deep Grammar

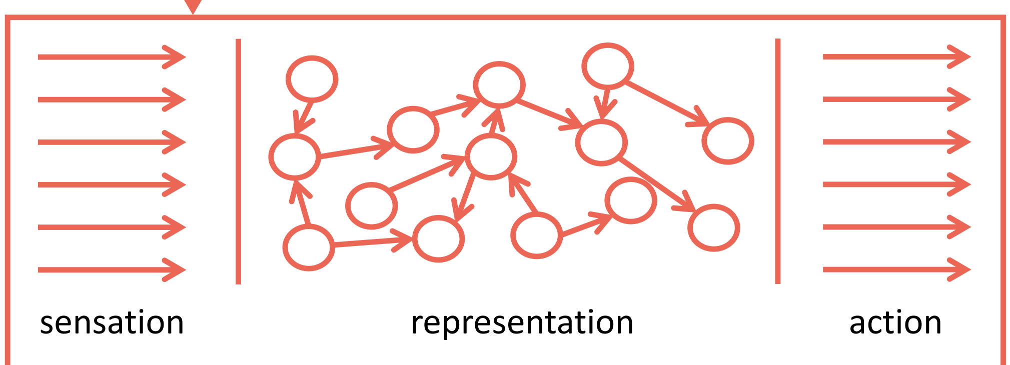
symbolic path

meaningless tokens

manual representations

world models

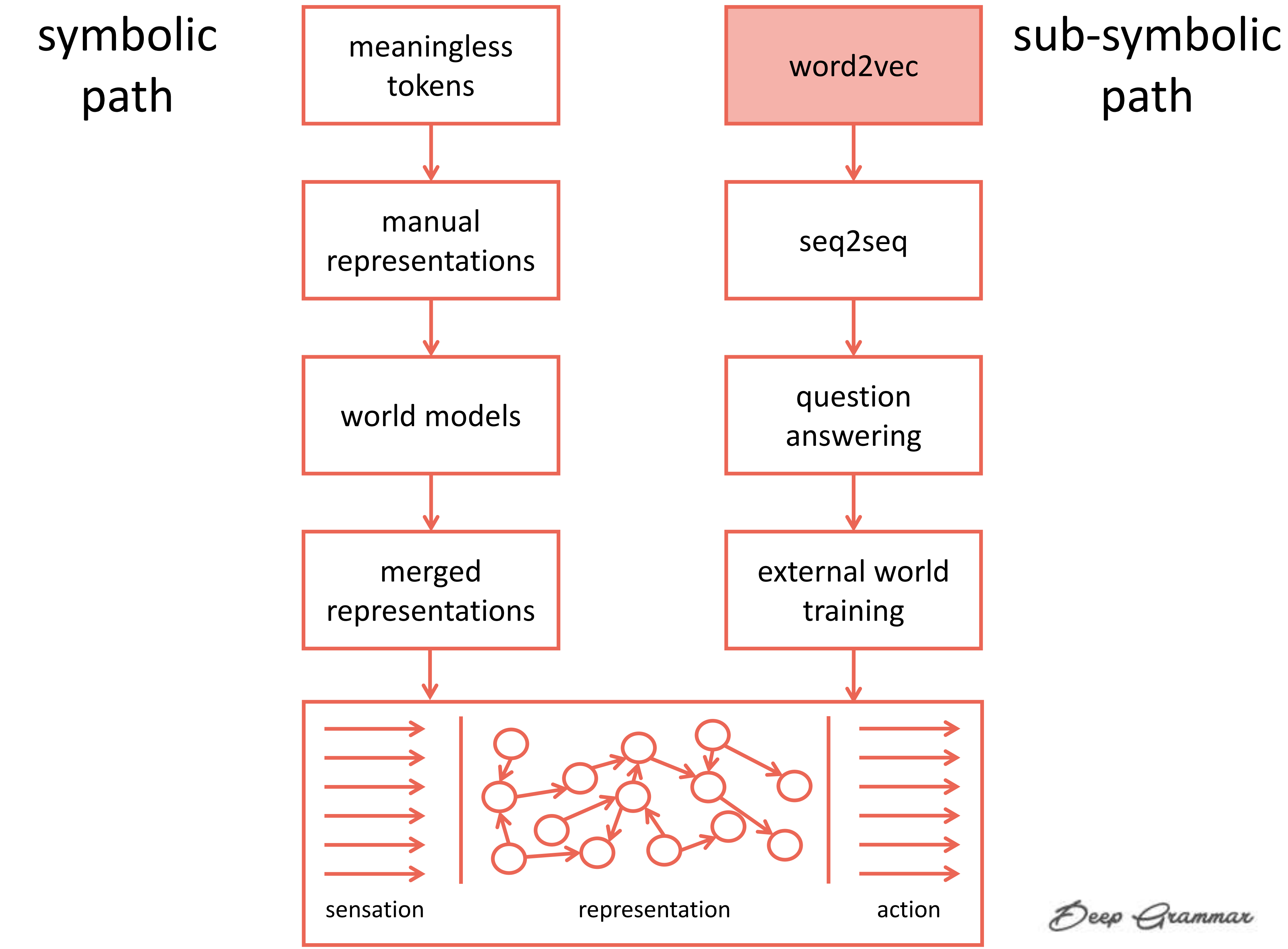
merged representations



sub-symbolic path

Neural networks (deep learning)
Great place to start is the landmark publication:

Parallel Distributed Processing, Vols. 1 and 2, Rumelhart and McClelland, 1987


symbolic path
meaningless tokens
manual representations
world models
merged representations
sub-symbolic path
word2vec
seq2seq
question answering
external world training
sensation
representation
action
Deep Grammar

word2vec

The word2vec model learns a vector for each word in the vocabulary.

The number of dimensions for each word vector is the same and is usually around 300.

Unlike the tf-idf vectors, word vectors are dense, meaning that most values are not 0.

word2vec

```
1. Initialize each word with a random vector
2. For each word w1 in the set of documents:
3.     For each word w2 around w1:
4.         Move vectors for w1 and w2 closer
           together and move all others and w1
           farther apart
5. Goto 2 if not done
```

- Skip-gram model [Mikolov et al., 2013]
- Note: there are really two vectors per word, because you don't want a word to be likely to be around itself, see Goldberg and Levyhttps://arxiv.org/pdf/1402.3722v1.pdf
- First saw that double-for-loop explanation from Christopher Moody

Deep Grammar

word2vec meaning

The quote we often see:

"You shall know a word by the company it keeps." J. R. Firth [1957]

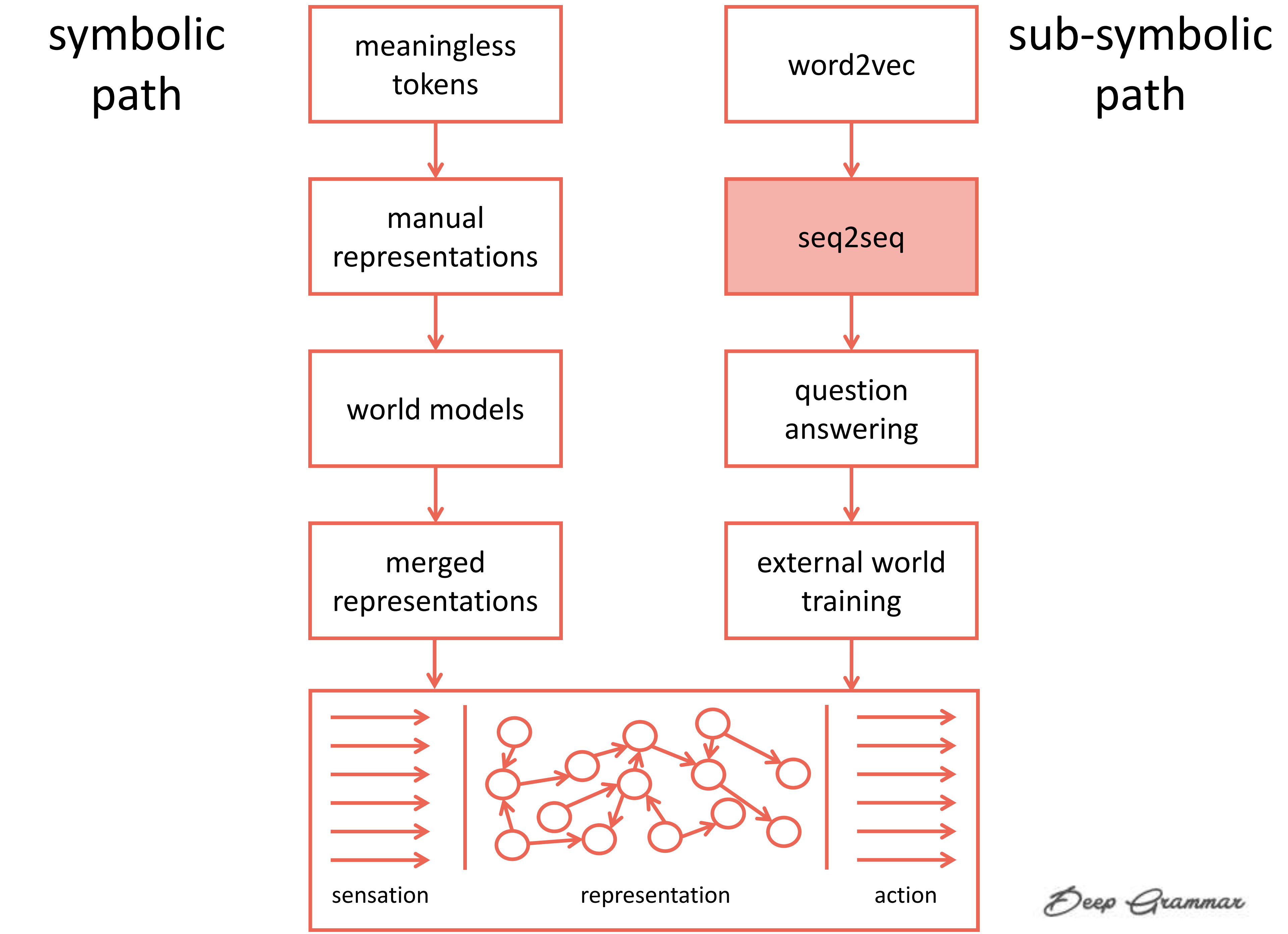
This seems at least kind of true.

- Vectors have internal structure [Mikolov et al., 2013]
- Italy – Rome = France – Paris
- King – Queen = Man – Woman

But ... words aren't grounded in experience; they are only grounded in being around other words.

Can also do word2vec on ConceptNet, see https://arxiv.org/pdf/1612.03975v1.pdf

Deep Grammar


symbolic path
meaningless tokens
manual representations
world models
merged representations
sub-symbolic path
word2vec
seq2seq
question answering
external world training
sensation
representation
action
Deep Grammar

seq2seq model

The seq2seq (sequence-to-sequence) model can encode sequences of tokens, such as sentences, into single vectors.

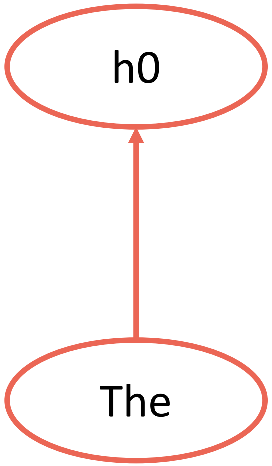
It can then decode these vectors into other sequences of tokens.

Both the encoding and decoding are done using recurrent neural networks (RNNs).

One obvious application for this is machine translation. For example, where the source sentences are English and the target sentences are Spanish.

Encoding sentence meaning into a vector

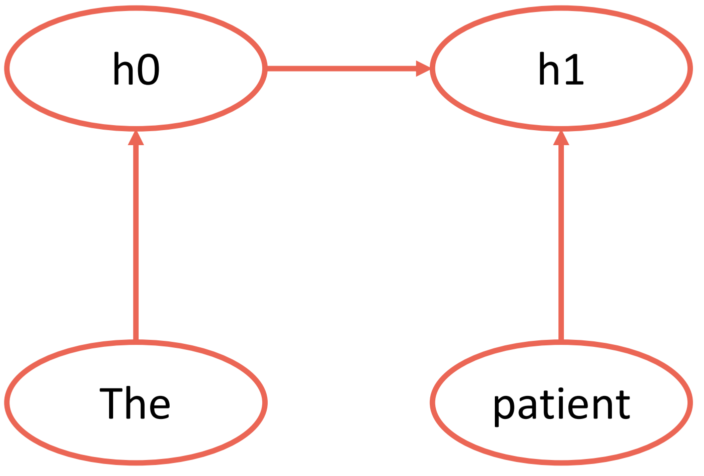
"The patient fell."



Deep Grammar

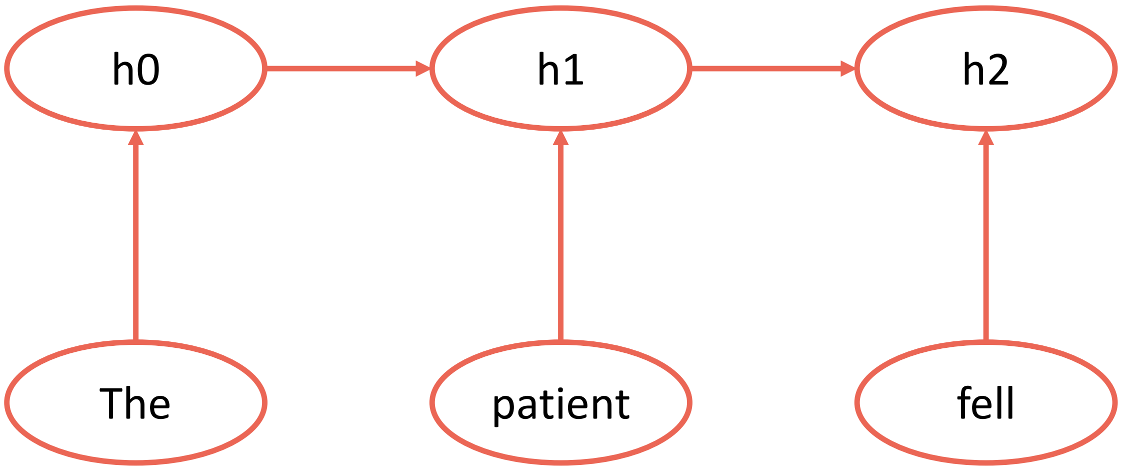
Encoding sentence meaning into a vector

"The patient fell."



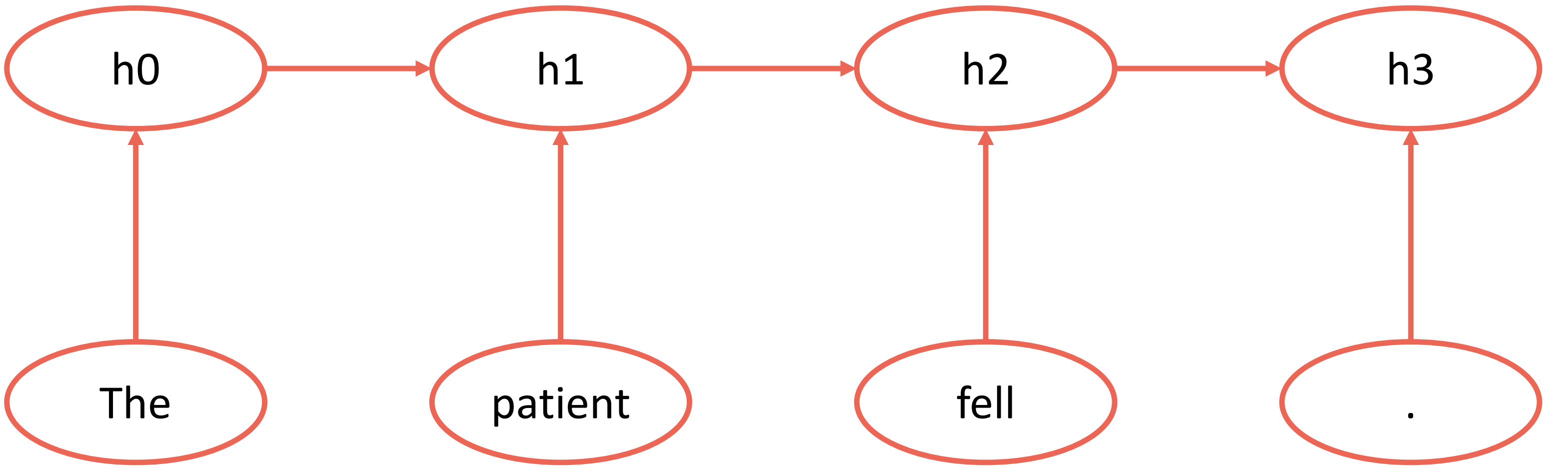
Encoding sentence meaning into a vector

"The patient fell."



Encoding sentence meaning into a vector

“The patient fell.”

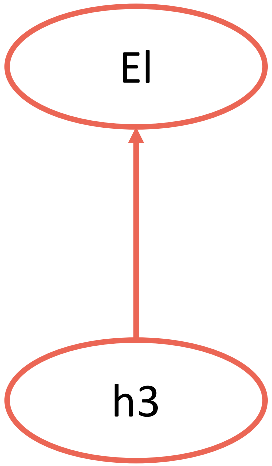


Like a hidden Markov model, but doesn’t make the Markov assumption and benefits from a vector representation.

Deep Grammar

Decoding sentence meaning

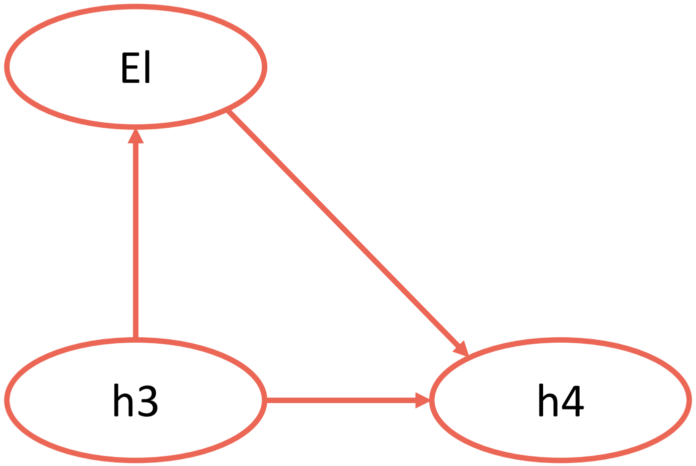
Machine translation, or structure learning more generally.



Deep Grammar

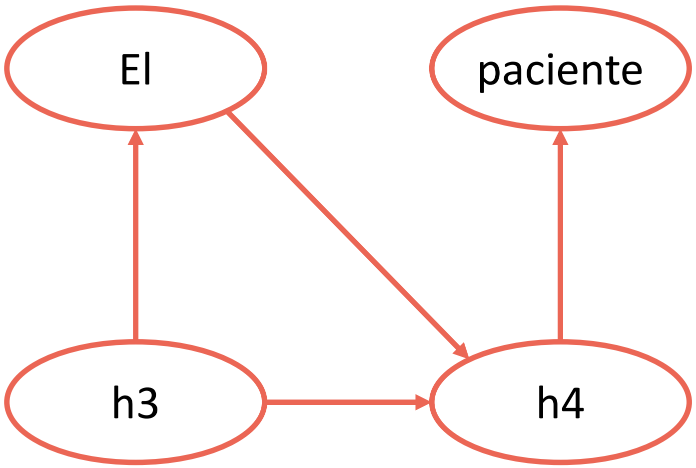
Decoding sentence meaning

Machine translation, or structure learning more generally.



Decoding sentence meaning

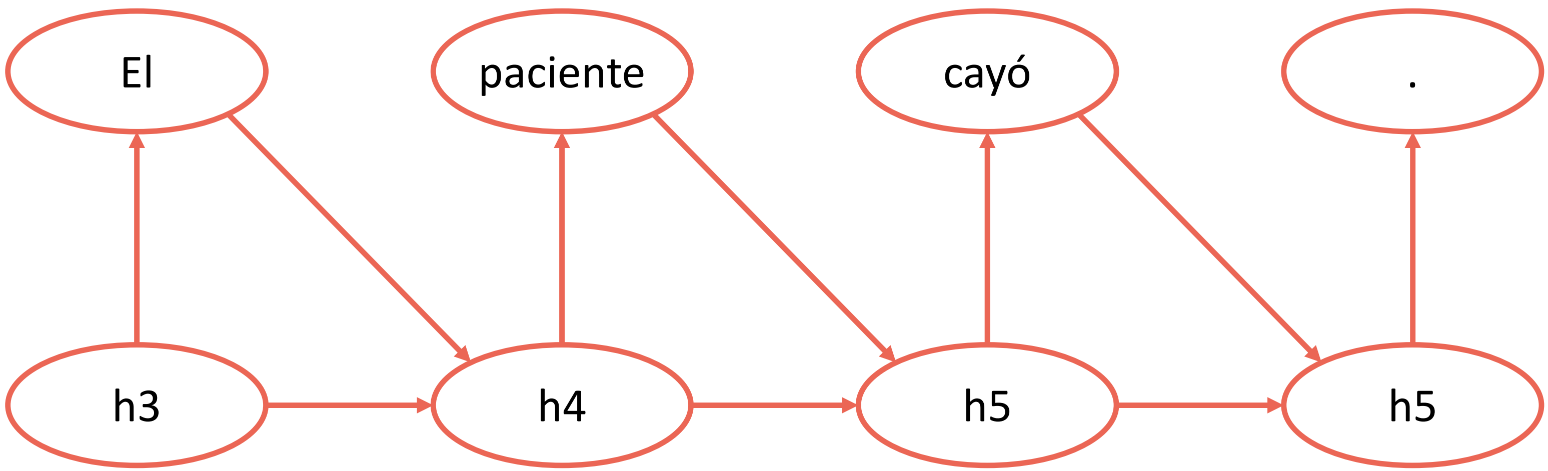
Machine translation, or structure learning more generally.



Deep Grammar

Decoding sentence meaning

Machine translation, or structure learning more generally.



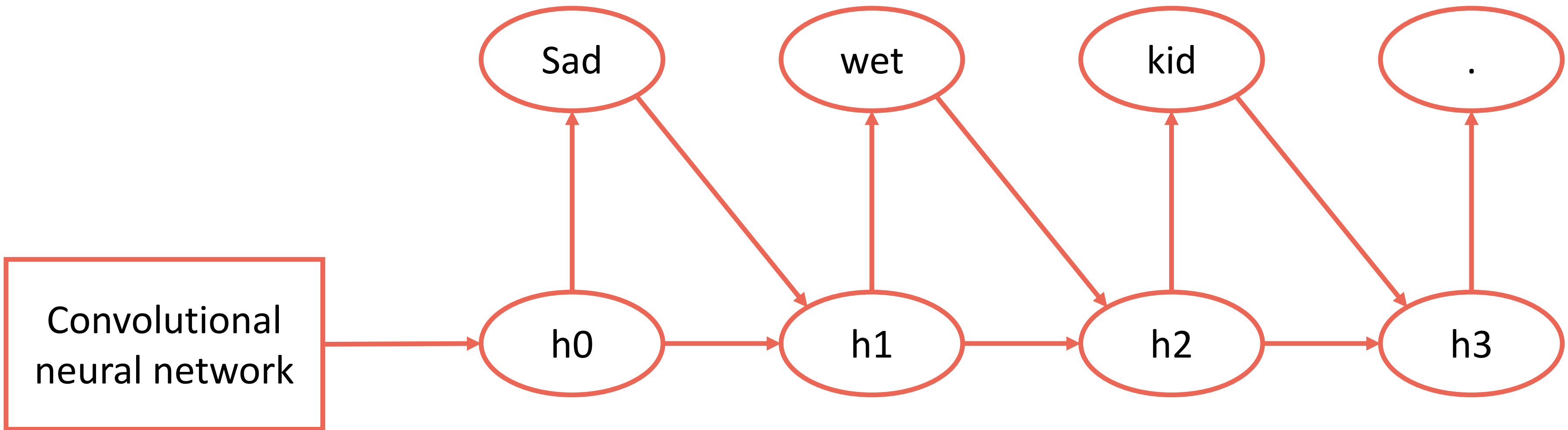
[Cho et al., 2014]

It keeps generating until it generates a stop symbol. Note that the lengths don't need to be the same. It could generate the correct "se cayó."

- Treats this task like it is devoid of meaning.
- Great that this can work on just about any kind of seq2seq problem, but this generality highlights its limitation for use as language understanding. No Chomsky universal grammar.

Deep Grammar

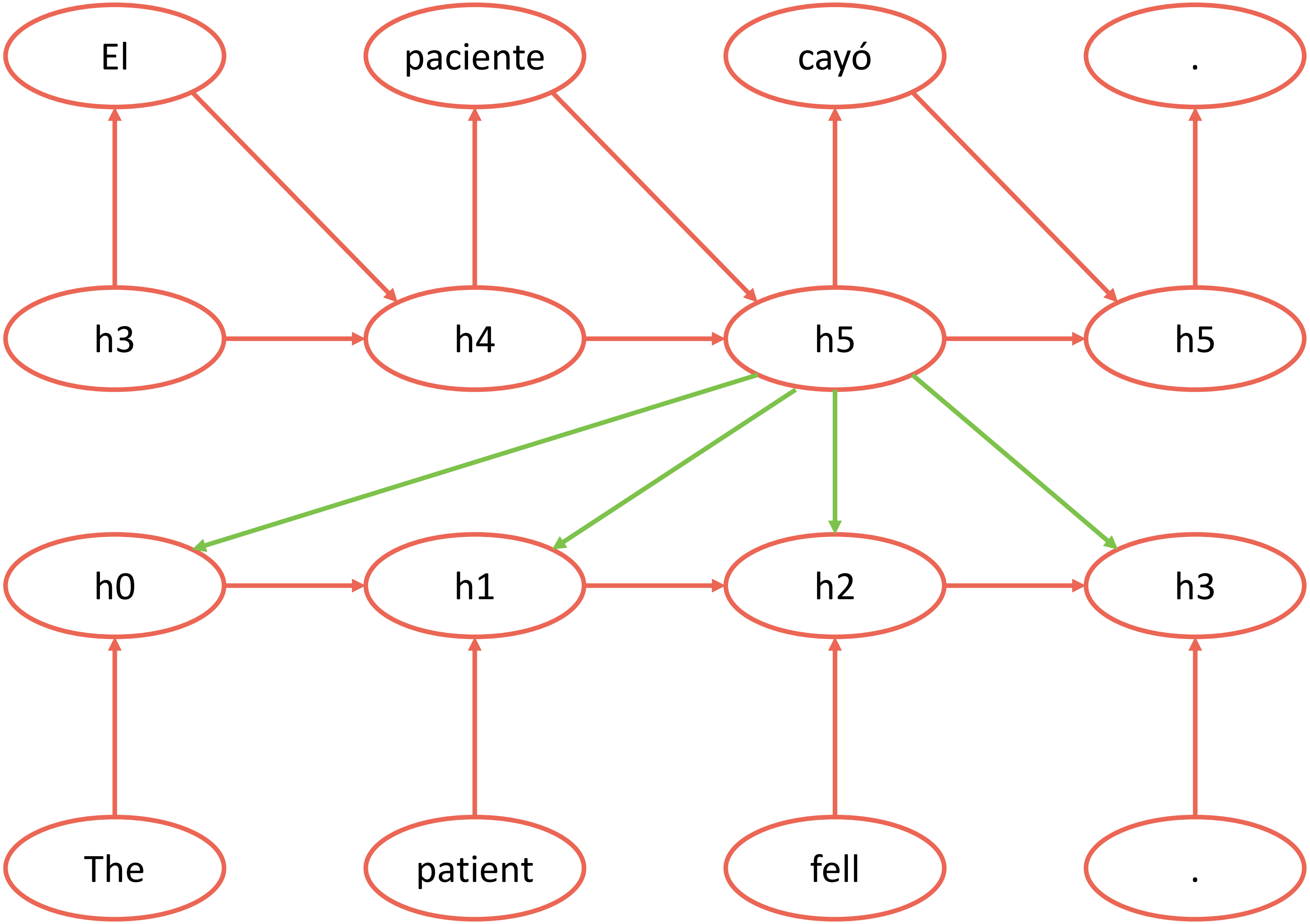
Generating image captions

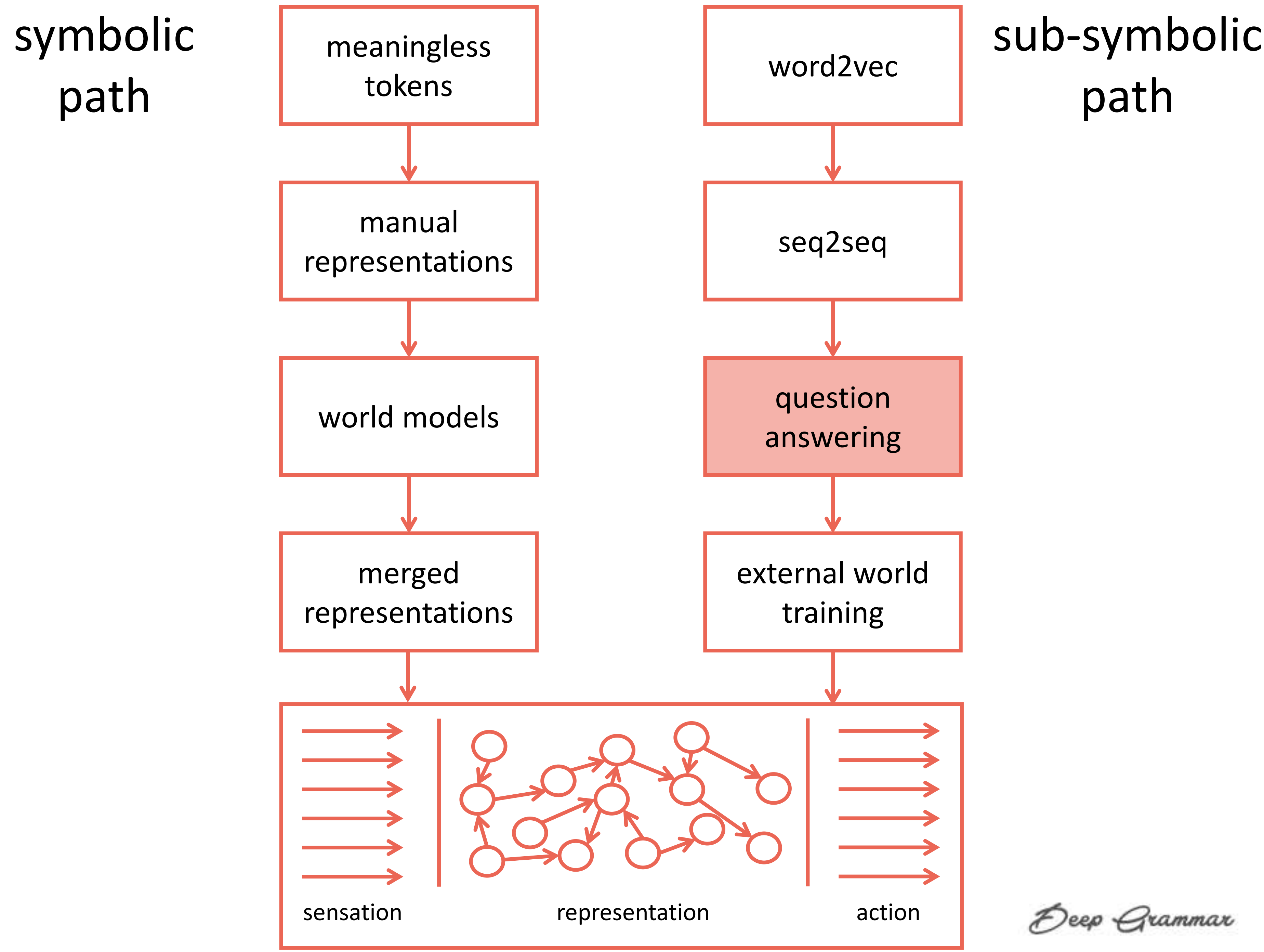


[Karpathy and Fei-Fei, 2015]
[Vinyals et al., 2015]

Deep Grammar

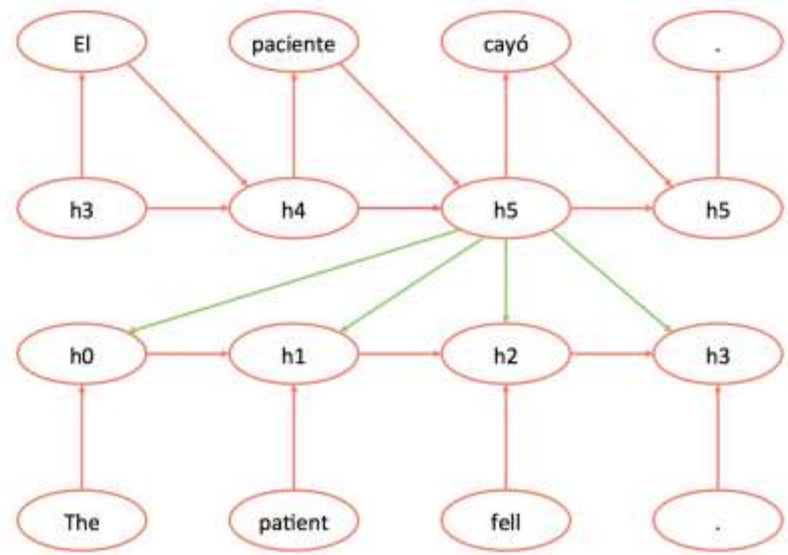
Attention [Bahdanau et al., 2014]




symbolic path
meaningless tokens
manual representations
world models
merged representations
sub-symbolic path
word2vec
seq2seq
question answering
external world training
sensation
representation
action
Deep Grammar

Deep learning and question answering

RNNs answer questions.



What is the translation of this phrase to French?

What is the next word?

Attention is useful for question answering.

This can be generalized to which facts the learner should pay attention to when answering questions.

Deep learning and question answering

Bob went home.
Tim went to the junkyard.
Bob picked up the jar.
Bob went to town.
Where is the jar? A: town

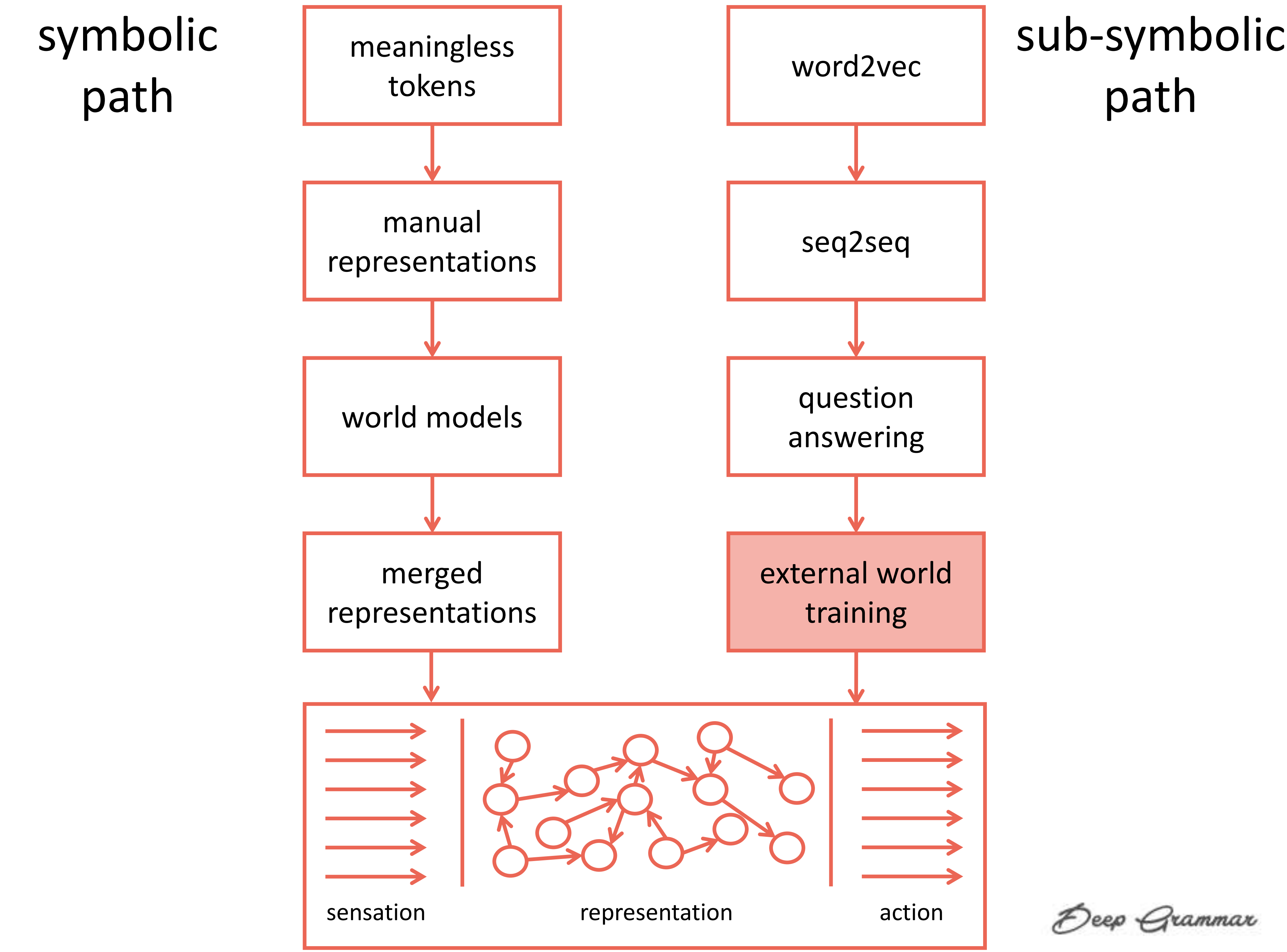
- Memory Networks [Weston et al., 2014]
- Updates memory vectors based on a question and finds the best one to give the output.

The office is north of the yard.
The bath is north of the office.
The yard is west of the kitchen.
How do you go from the office to the kitchen? A: south, east

- Neural Reasoner [Peng et al., 2015]
- Encodes the question and facts in many layers, and the final layer is put through a function that gives the answer.

Deep learning and question answering

The network is learning linkages between sequences of symbols, but these kinds of stories do not have sufficiently rich linkages to our world.

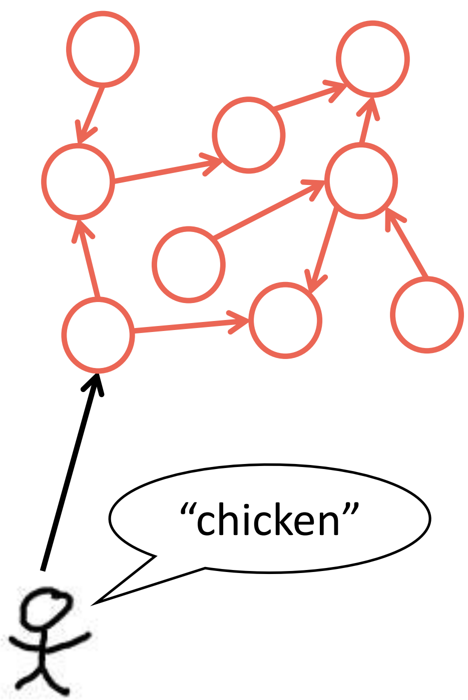

symbolic path
meaningless tokens
manual representations
world models
merged representations
word2vec
seq2seq
question answering
external world training
sub-symbolic path
sensation
representation
action
Deep Grammar

External world training

If we want to talk to machines

1. We need to train them in an environment as much like our own as possible
2. Can't just be dialog!

   Harnad [1990] http://users.ecs.soton.ac.uk/harnad/Papers/Harnad/harnad90.sgproblem.html



To understand "chicken" we need the machine to have had as much experience with chickens as possible.

When we say "chicken" we don't just mean the bird, we mean everything one can do with it and everything it represents in our culture.

There has been work in this direction

Industry

- OpenAI
  - Universe: train on screens with VNC
  - Now with Grand Theft Auto! https://openai.com/blog/GTA-V-plus-Universe/
- Google
  - Mikolov et al., *A Roadmap towards Machine Intelligence.* They define an artificial environment. https://arxiv.org/pdf/1511.08130v2.pdf
- Facebook
  - Weston, memory networks to dialogs https://arxiv.org/pdf/1604.06045v7.pdf
  - Kiela et al., *Virtual Embodiment: A Scalable Long-Term Strategy for Artificial Intelligence Res*. Advocate using video games "with a purpose." https://arxiv.org/pdf/1610.07432v1.pdf

Academia

- Ray Mooney
  - Maps text to situations http://videolectures.net/aaai2013_mooney_language_learning/
- Luc Steels
  - Robots come up with vocabulary and simple grammar
- Narasimhan et al.
  - Train a neural net to play text-based adventure games https://arxiv.org/pdf/1506.08941v2.pdf

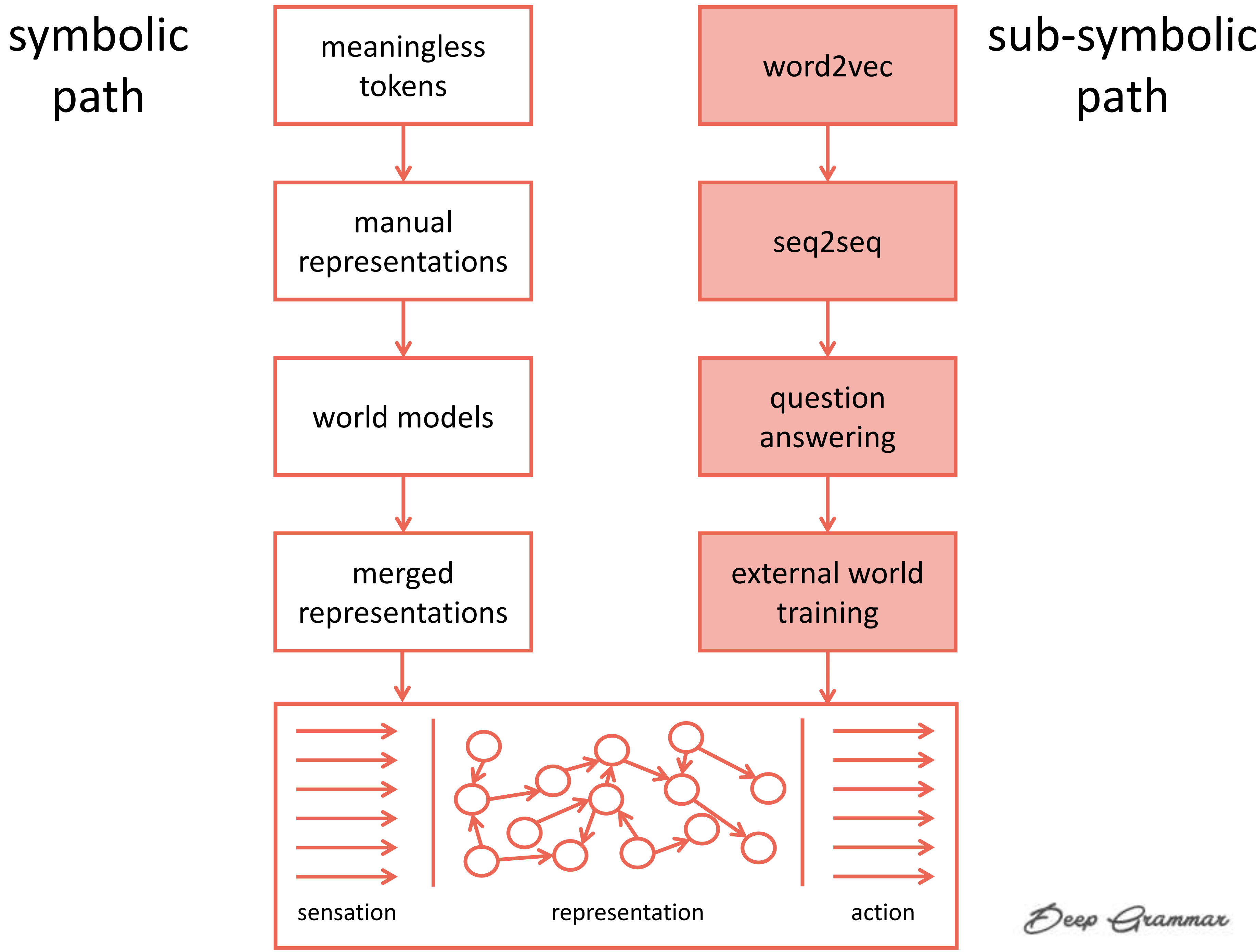
iCub

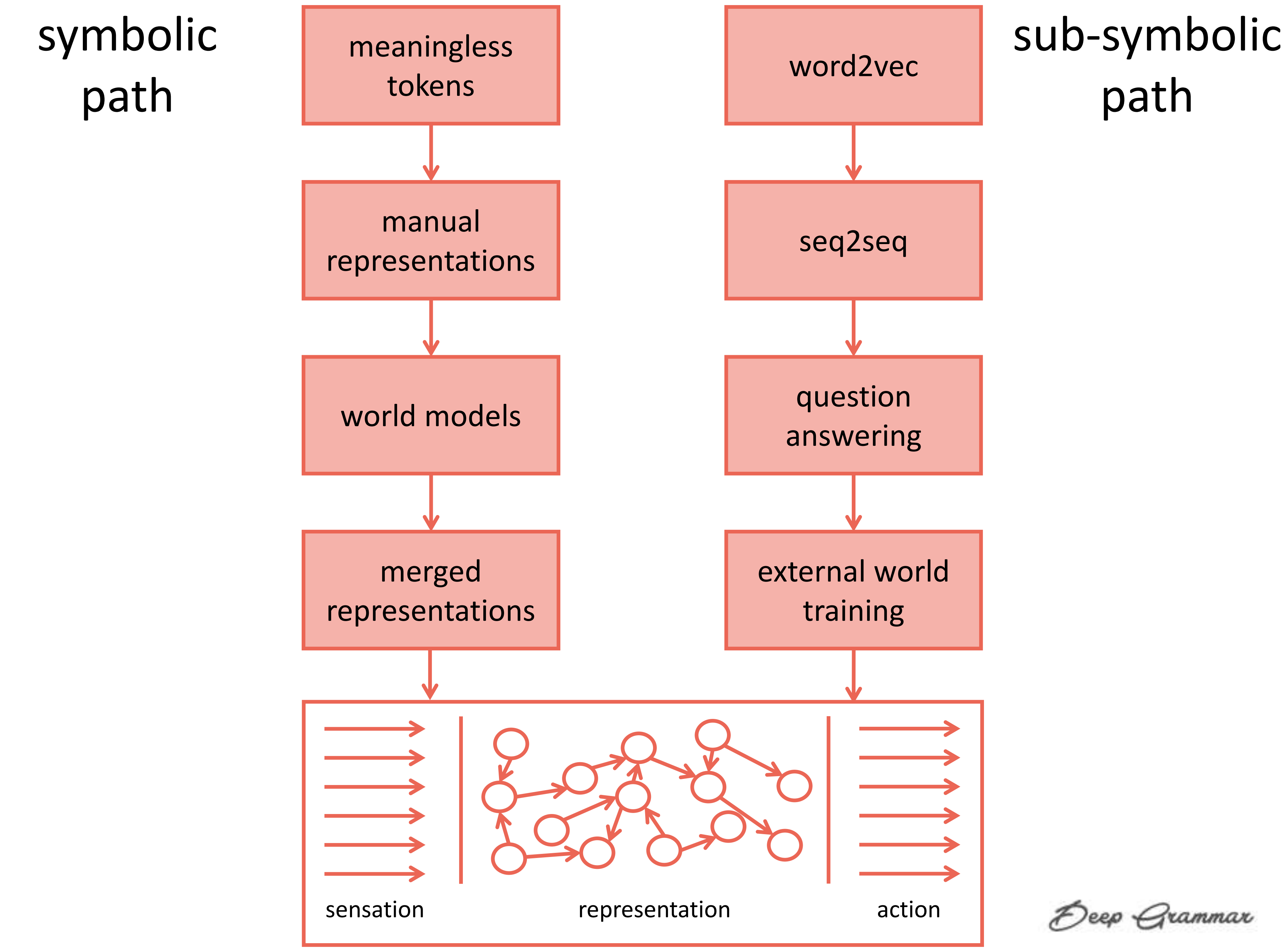
But we need more training centered in our world

Maybe if Amazon Alexa had a camera and rotating head?

How far could we get without the benefit of a teacher?

- Could we use eye gaze as a cue? [Yu and Ballard, 2010]


symbolic
path
meaningless
tokens
manual
representations
world models
merged
representations
word2vec
seq2seq
question
answering
external world
training
sub-symbolic
path
sensation
representation
action
Deep Grammar


symbolic path
meaningless tokens
manual representations
world models
merged representations
sub-symbolic path
word2vec
seq2seq
question answering
external world training
sensation
representation
action
Deep Grammar

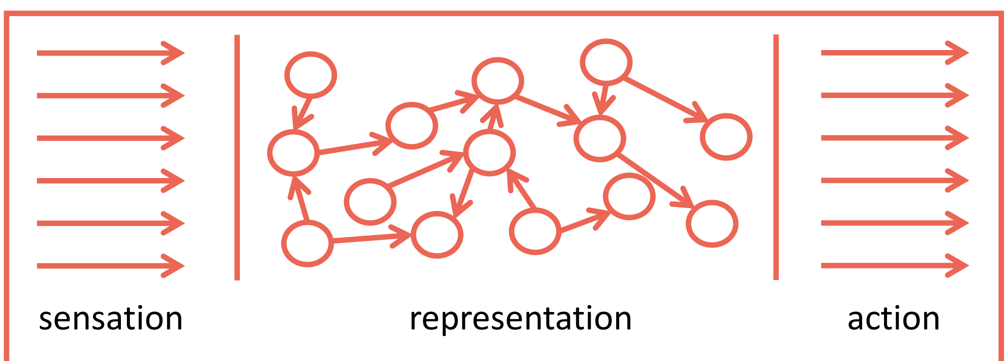
Two paths from NLP to AI

symbolic path

Building world models based on large, deep, and organized representations

sub-symbolic path

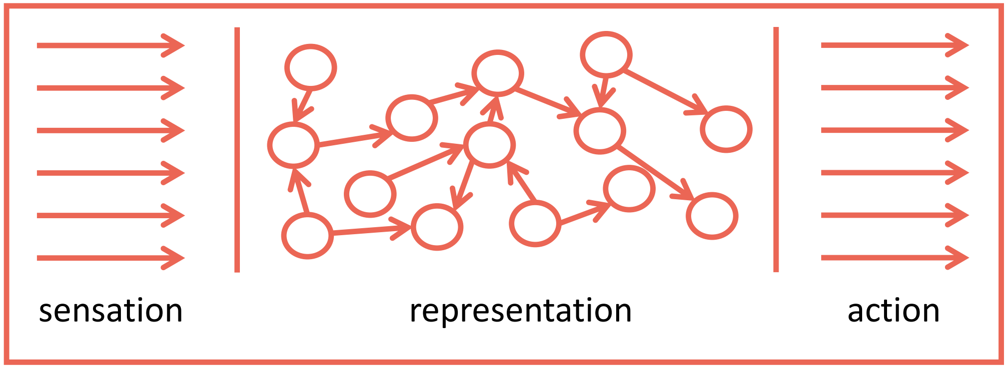
Training large neural networks in an environment with similar objects, relationships, and dynamics as our own



Deep Grammar

Final thought

Open problem: What is the simplest commercially viable task that requires commonsense knowledge and reasoning?



Deep Grammar

Deep Grammar

Thanks for listening

Jonathan Mugan

@jmugan

www.deepgrammar.com