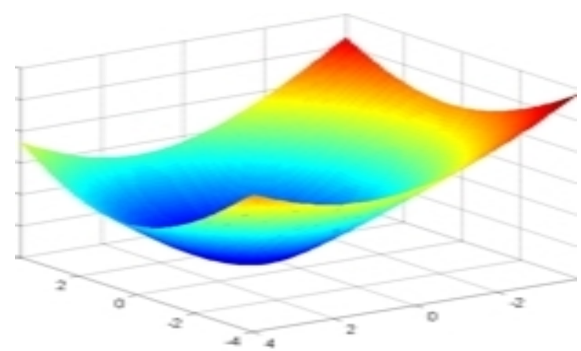
2. What's Deep Learning (DL)?

- **Deep learning** is a subfield of **machine learning**

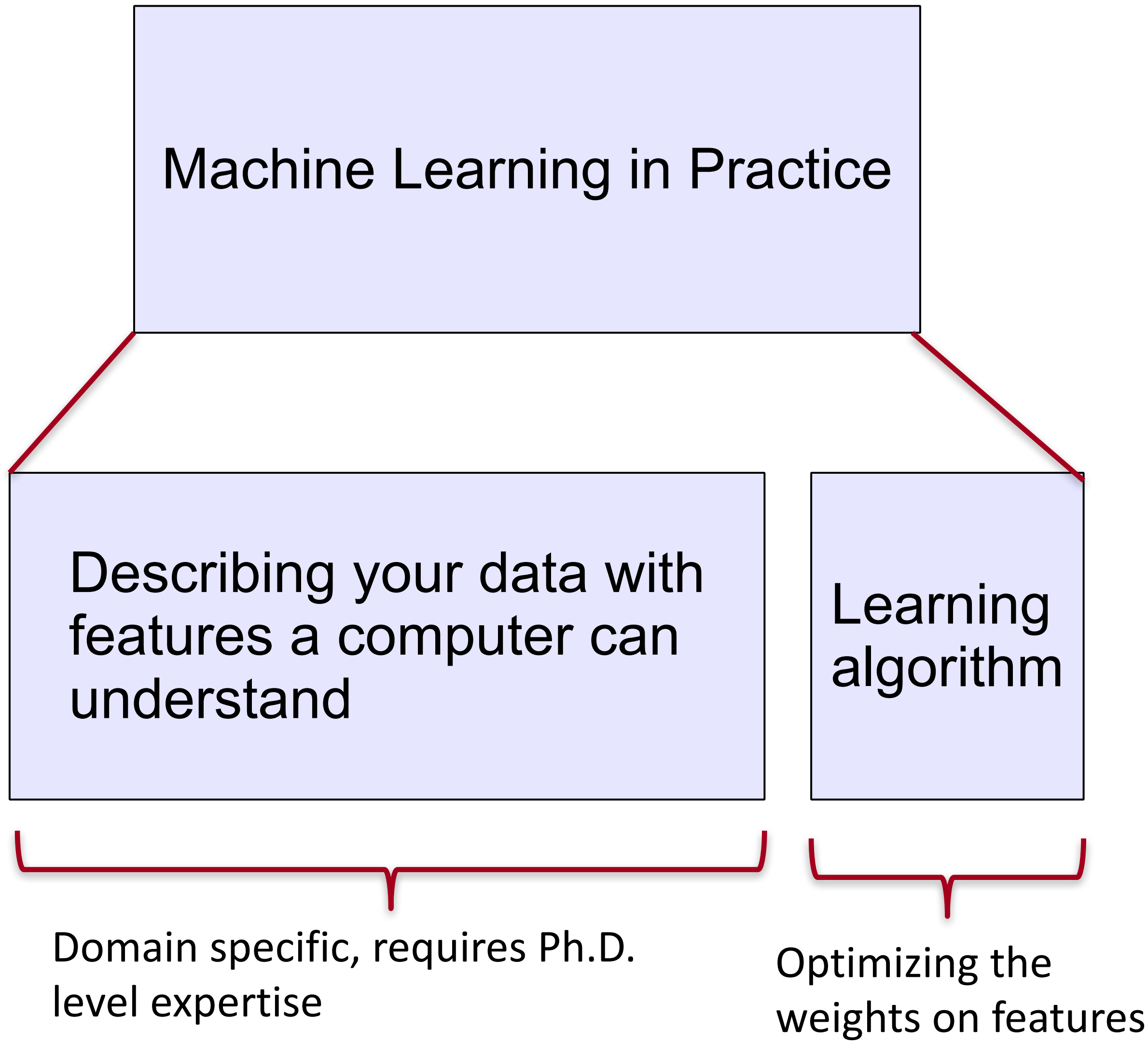
- Most machine learning methods work well because of **human-designed representations** and **input features**
  - For example: features for finding named entities like locations or organization names (Finkel et al., 2010):

Feature	NER
Current Word	✓
Previous Word	✓
Next Word	✓
Current Word Character n-gram	all
Current POS Tag	✓
Surrounding POS Tag Sequence	✓
Current Word Shape	✓
Surrounding Word Shape Sequence	✓
Presence of Word in Left Window	size 4
Presence of Word in Right Window	size 4

- Machine learning becomes just optimizing weights to best make a final prediction

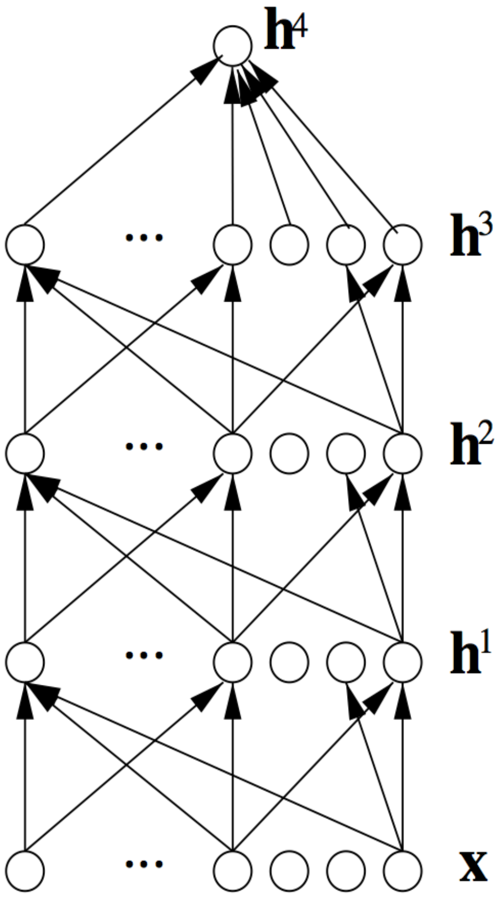


Machine Learning vs. Deep Learning



What's Deep Learning (DL)?

- In contrast to standard machine learning,
- Representation learning attempts to automatically learn good features or representations
- Deep learning algorithms attempt to learn (multiple levels of) representations (here: h^1,h^2,h^3) and an output (h^4)
- From "raw" inputs **x** (e.g. sound, pixels, characters, or words)



On the history of "Deep Learning"

- We will focus on different kinds of **neural networks**
- The dominant model family inside deep learning

- Only clever terminology for stacked logistic regression units?
  - Maybe, but interesting modeling principles (end-to-end) and actual connections to neuroscience in some cases.
  - Recently: Differentiable Programming – becomes clear later

- We will not take a historical approach but instead focus on methods which work well on NLP problems now
- For a long history of deep learning models (starting ~1960s), see: [Deep Learning in Neural Networks: An Overview]() by Jürgen Schmidhuber

Reasons for Exploring Deep Learning

- Manually designed features are often over-specified, incomplete and take a long time to design and validate

- **Learned Features** are easy to adapt, fast to learn

- Deep learning provides a very flexible, (almost?) universal, learnable framework for **representing** world, visual and linguistic information.

- Deep learning can learn **unsupervised** (from raw text) and **supervised** (with specific labels like positive/negative)

Reasons for Exploring Deep Learning

- In ~2010 **deep** learning techniques started outperforming other machine learning techniques. Why this decade?

- Large amounts of training data favor deep learning
- Faster machines and multicore CPU/GPUs favor Deep Learning
- New models, algorithms, ideas
  - Better, more flexible learning of intermediate representations
  - Effective end-to-end joint system learning
  - Effective learning methods for using contexts and transferring between tasks
  - Better regularization and optimization methods

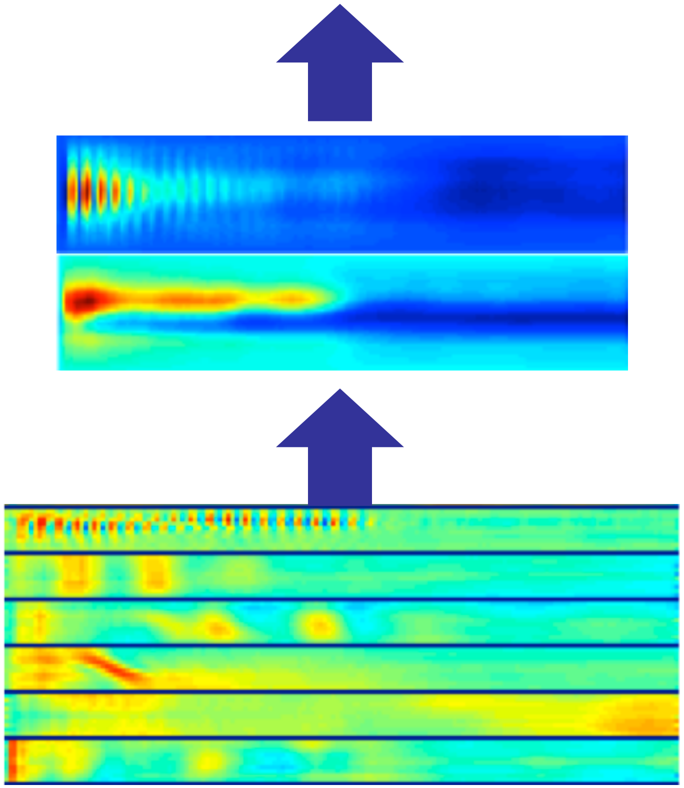
→ **Improved performance** (first in speech and vision, then NLP)

Deep Learning for Speech

- The first breakthrough results of "deep learning" on large datasets happened in speech recognition
- Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition Dahl et al. (2010)

Acoustic model and WER	RT03S FSH	Hub5 SWB
Traditional features	**27.4**	**23.6**
Deep Learning	**18.5** (−33%)	**16.1** (−32%)

Phonemes/Words



Deep Learning for Computer Vision

First major focus of deep learning groups was computer vision

The breakthrough DL paper: ImageNet Classification with Deep Convolutional Neural Networks by Krizhevsky, Sutskever, & Hinton, 2012, U. Toronto. 37% error red.

Zeiler and Fergus (2013)

5. Deep NLP = Deep Learning + NLP

Combine ideas and goals of NLP with using representation learning and deep learning methods to solve them

Several big improvements in recent years in NLP

- **Linguistic levels**: (speech), words, syntax, semantics
- **Intermediate tasks/tools:** parts-of-speech, entities, parsing
- **Full applications**: sentiment analysis, question answering, dialogue agents, machine translation

Word meaning as a neural word vector – visualization

$$expect = \begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \\ 0.487 \end{pmatrix}$$

Word similarities

Nearest words to frog:

1. frogs
2. toad
3. litoria
4. leptodactylidae
5. rana
6. lizard
7. eleutherodactylus

litoria

leptodactylidae

rana

eleutherodactylus

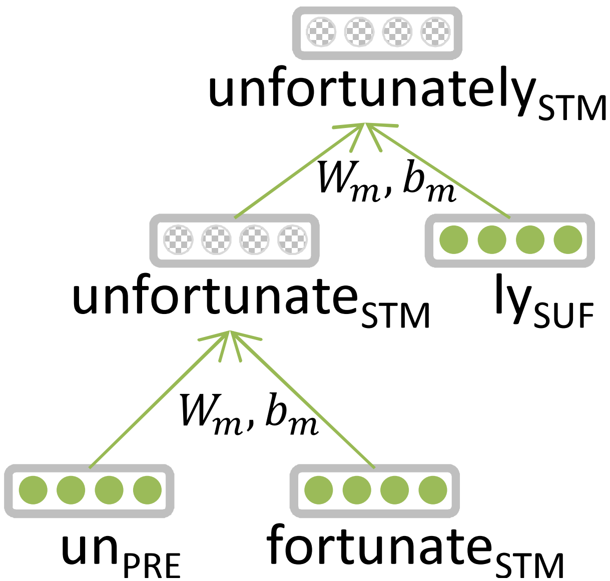
http://nlp.stanford.edu/projects/glove/

Representations of NLP Levels: Morphology

- Traditional: Words are made of morphemes

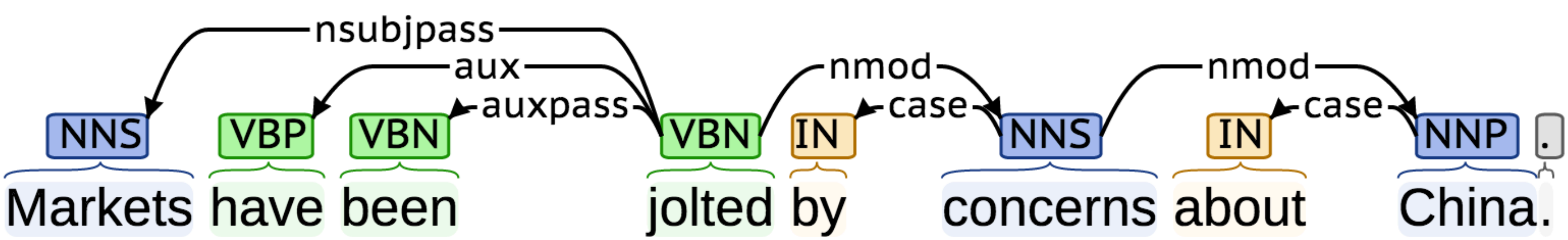
prefix	stem	suffix
un	interest	ed

- DL:
  - every morpheme is a vector
  - a neural network combines two vectors into one vector
  - Luong et al. 2013



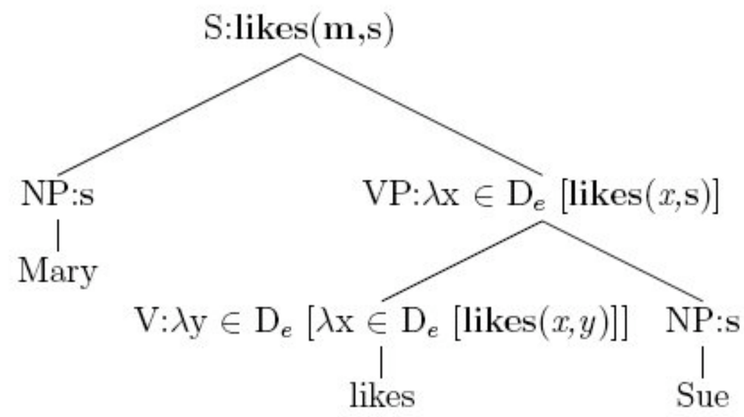
NLP Tools: Parsing for sentence structure

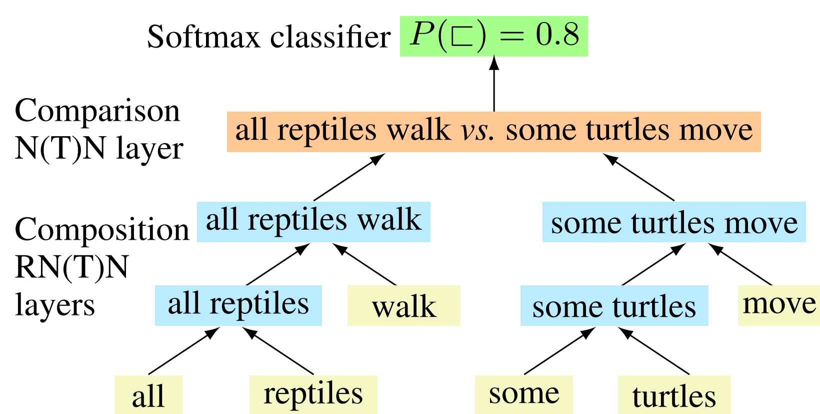
- Neural networks can accurately determine the grammatical structure of sentences
- This supports interpretation and may help in disambiguation



Representations of NLP Levels: Semantics

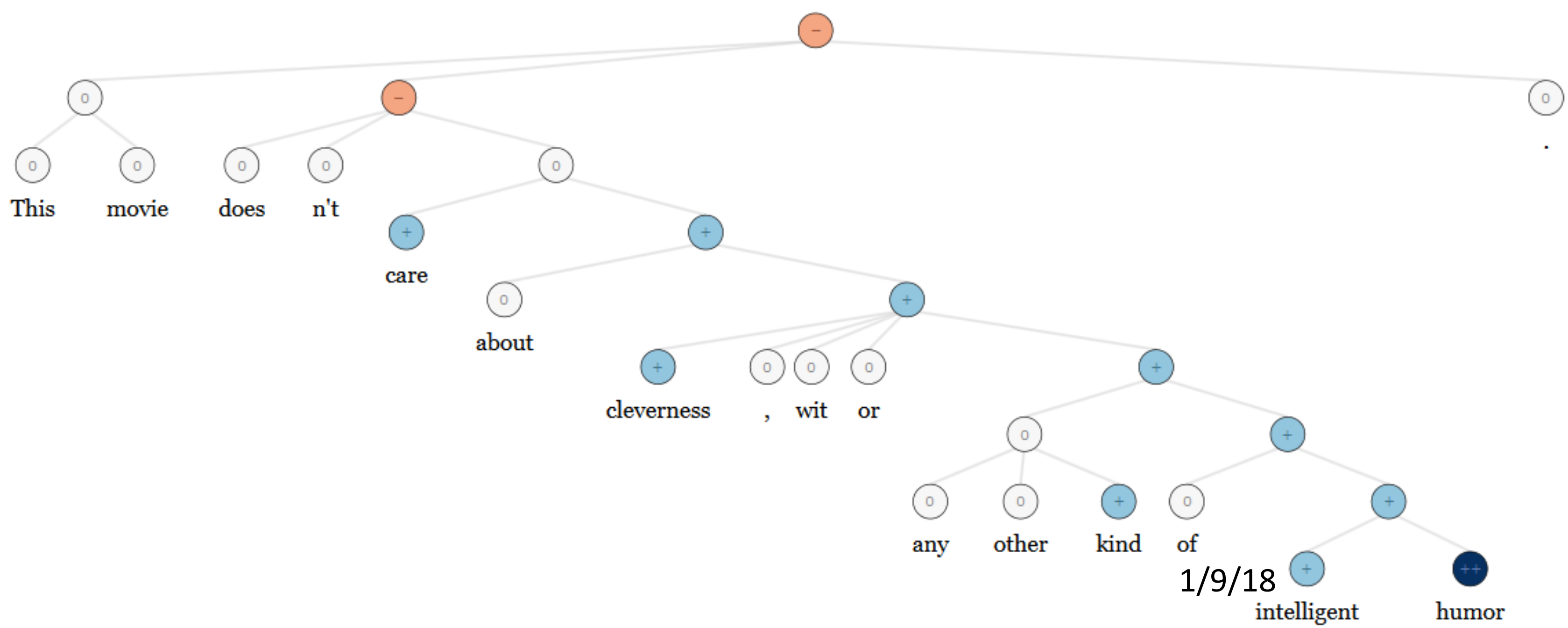
- Traditional: Lambda calculus
  - Carefully engineered functions
  - Take as inputs specific other functions
  - No notion of similarity or fuzziness of language
- DL:
  - Every word and every phrase and every logical expression is a vector
  - a neural network combines two vectors into one vector
  - Bowman et al. 2014





NLP Applications: Sentiment Analysis

- Traditional: Treat sentence as a bag-of-words (ignore word order); consult a curated list of "positive" and "negative" words to determine sentiment of sentence. Need hand-designed features to capture negation! --> Ain’t gonna capture everything 😖
- Same deep learning model that could be used for morphology, syntax and logical semantics → RecursiveNN (aka TreeRNNs)



Question Answering

- Traditional: A lot of feature engineering to capture world and other knowledge, e.g., regular expressions, Berant et al. (2014)

Is main verb trigger?

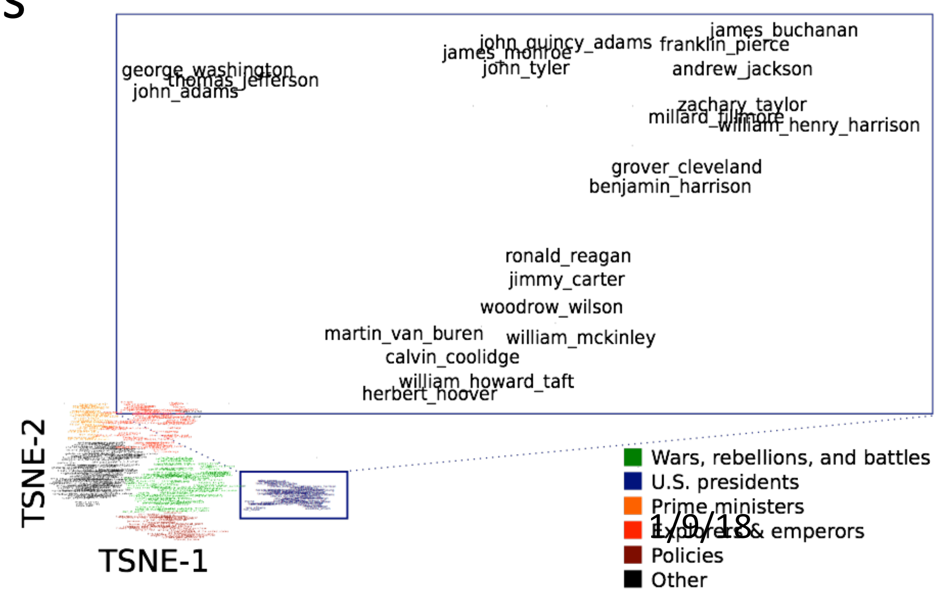
Yes

Condition	Regular Exp.
Wh- word subjective?	AGENT
Wh- word object?	THEME

No

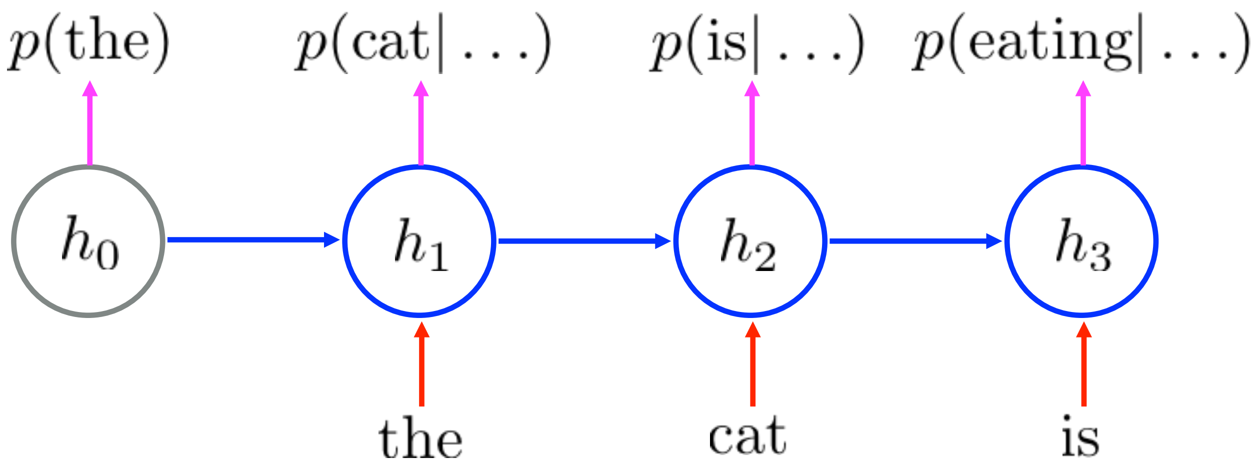
Condition	Regular Exp.
default	(ENABLE|SUPER)$^{+}$
DIRECT	(ENABLE|SUPER)
PREVENT	(ENABLE|SUPER)* PREVENT(ENABLE|SUPER)*

- DL: Again, a deep learning architecture can be used!
- Facts are stored in vectors



Dialogue agents / Response Generation

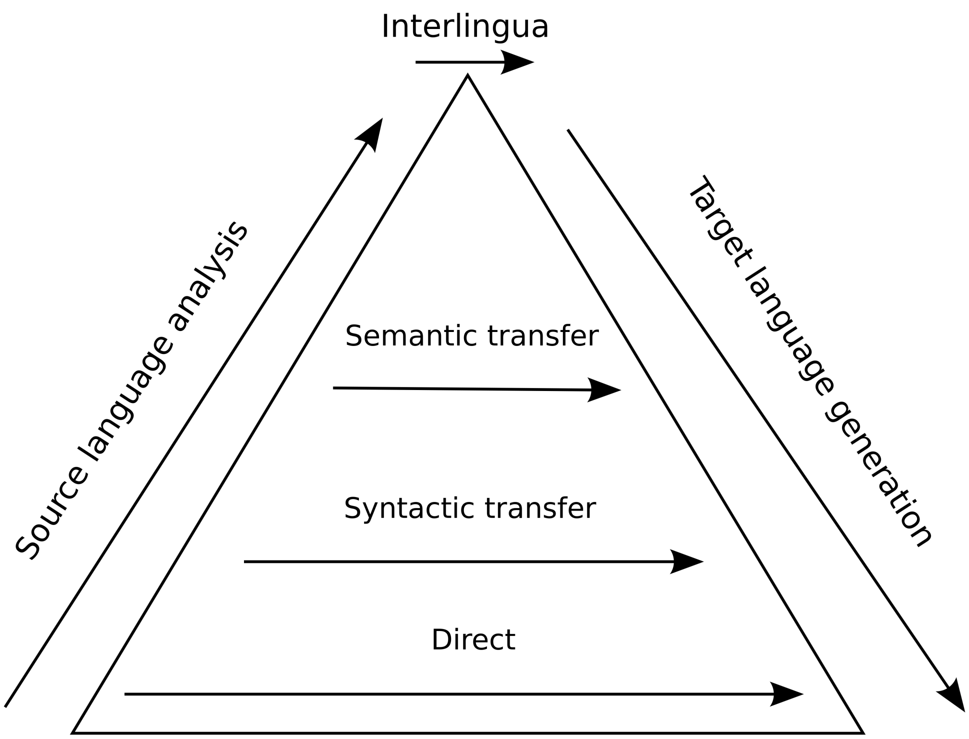
- A simple, successful example is the auto-replies available in the Google Inbox app
- An application of the powerful, general technique of **Neural Language Models**, which are an instance of Recurrent Neural Networks



Machine Translation

- Many levels of translation have been tried in the past:

- Traditional MT systems are very large complex systems

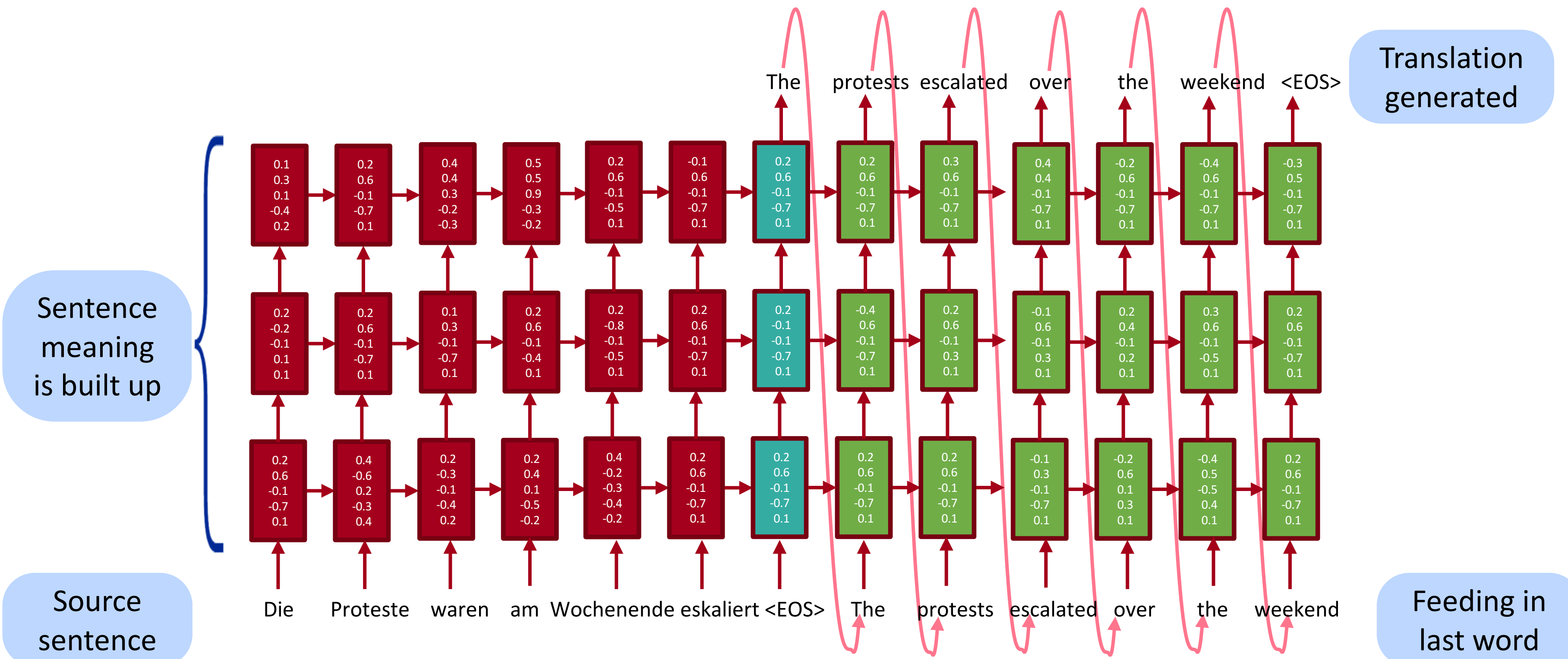


- What do you think is the interlingua for the DL approach to translation?

1/9/18

Neural Machine Translation

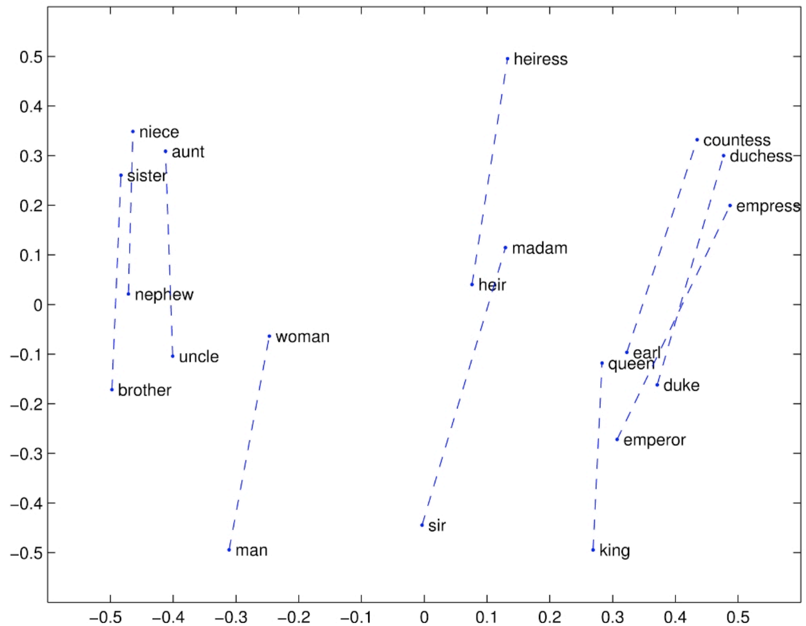
Source sentence is mapped to **vector**, then output sentence generated
[Sutskever et al. 2014, Bahdanau et al. 2014, Luong and Manning 2016]



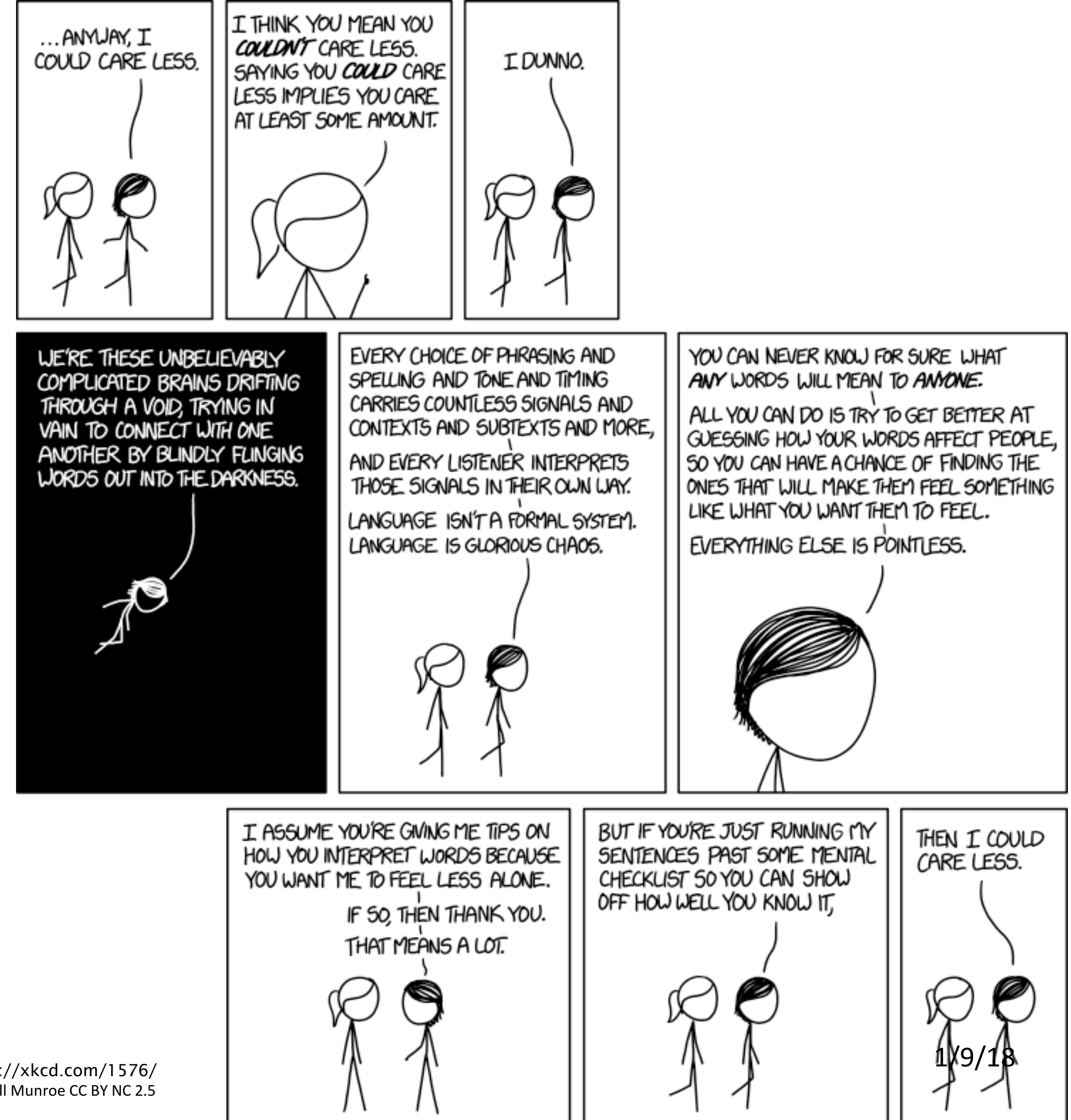
Now live for some languages in Google Translate (etc.), with big error reductions!

Conclusion: Representation for all levels? Vectors

We will study in the next lecture how we can learn vector representations for words and what they actually **represent.**



Next week: how neural networks work and how they can use these vectors for all NLP levels and many different applications



https://xkcd.com/1576/