

David Mareček, Rudolf Rosa  
marecek@ufal.mff.cuni.cz, rosa@ufal.mff.cuni.cz

# From Balustrades to Pierre Vinken:

## Looking for Syntax in Transformer Self-Attentions

Charles University, Prague  
Faculty of Mathematics and Physics  
Institute of Formal and Applied Linguistics

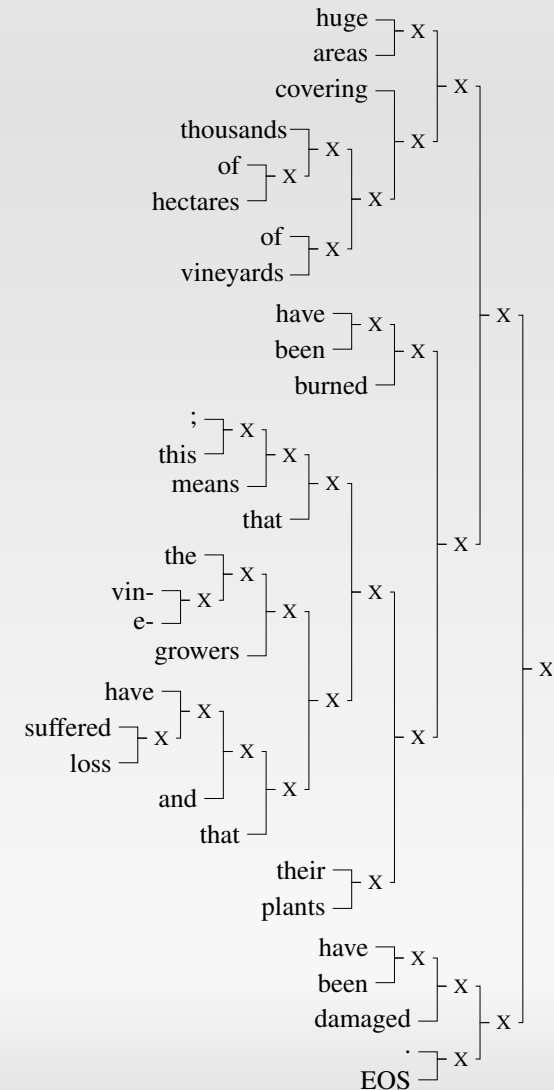
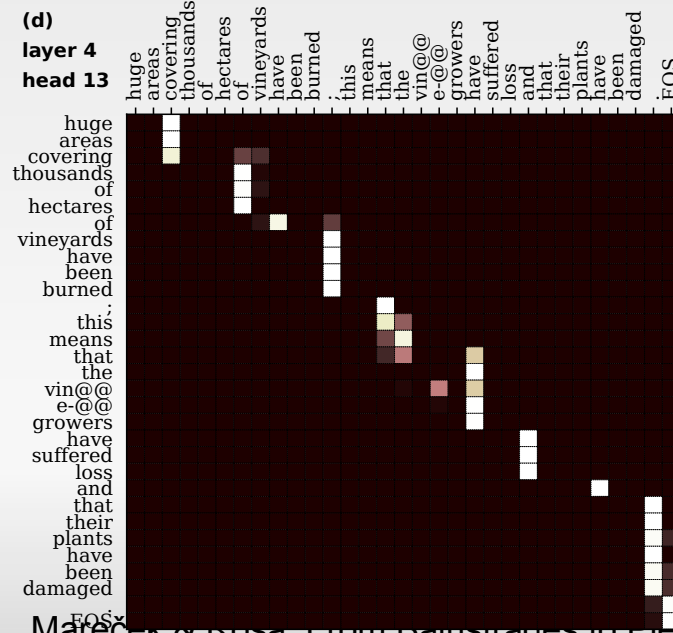
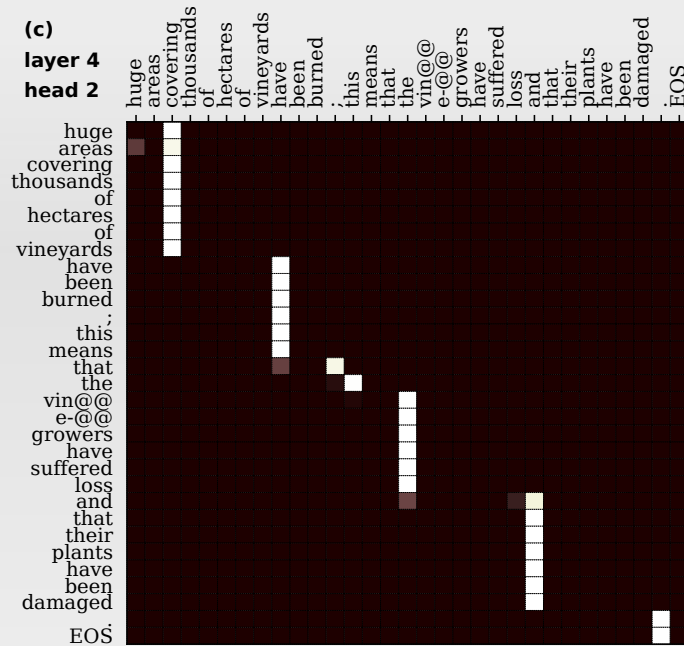


BlackboxNLP Workshop, Firenze, 1 August 2019

# From balustrades to Pierre Vinken

- fotka balustrády na schodišti
- fotka Pierra Vinkena

# Transformer self-att. → syntax



# Overview

- Experiment setup
- Inspection of self-attention heads
- Extracting and scoring phrase candidates
- Linguistically uninformed CKY parsing
- Results
- Summary

# Experiment setup

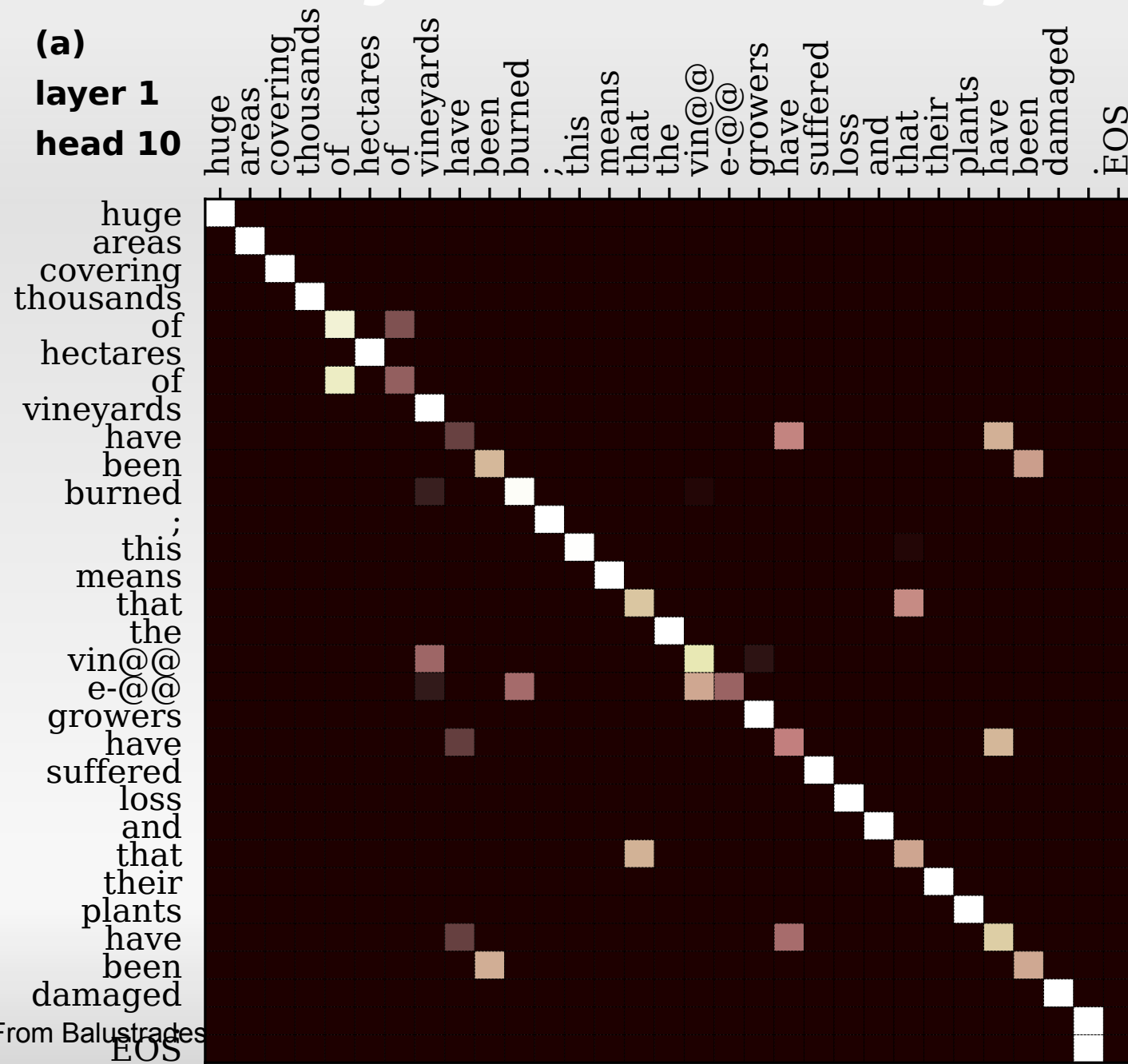
- transformer neural machine translation encoder
  - 6 layers x 16 heads, 100k shared BPEs...
- 6 language pairs: fr  $\leftrightarrow$  en, de  $\leftrightarrow$  en, fr  $\leftrightarrow$  de
  - Europarl training data
- analyze encoder self-attention matrices
- extract constituency syntax trees
- compare against Stanford parser syntax trees
  - trained on linguistically annotated treebanks:  
Penn Treebank, Negra Corpus, French Treebank

# Transformer NMT

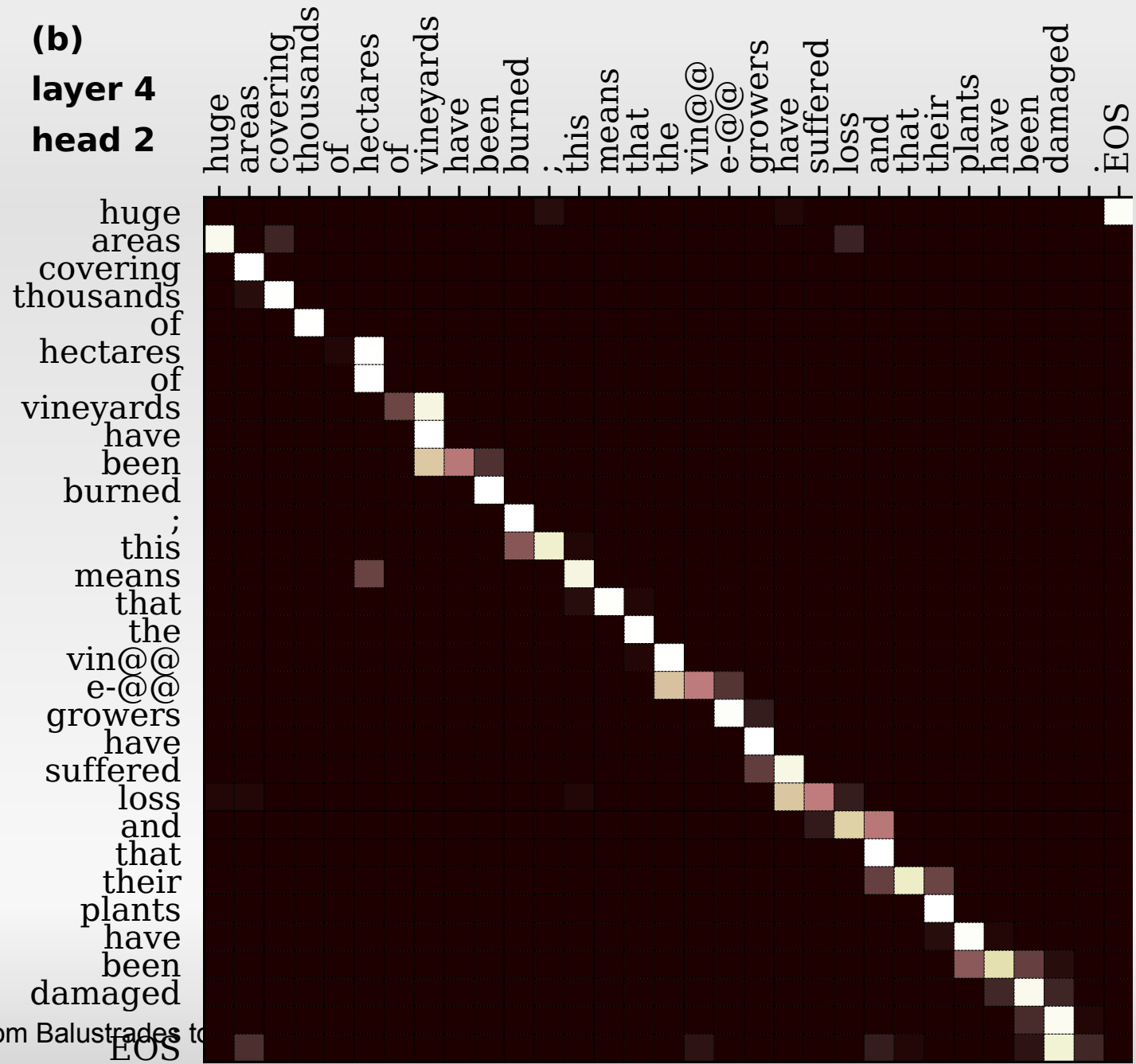
- tady asi strukturu transformera trochu
- at' je jasny odkud' taham ty self attention matrices
- positions ~ input words (actually subwords)
- each head attends to some words...
- one example sentence throughout all slides
  - Huge areas covering thousands of hectares of vineyards have been burned; this means that the vine-growers have suffered loss and that their plants have been damaged.

# Diagonal (word identity)

[SVG kvalitnější ale hnusnej font]

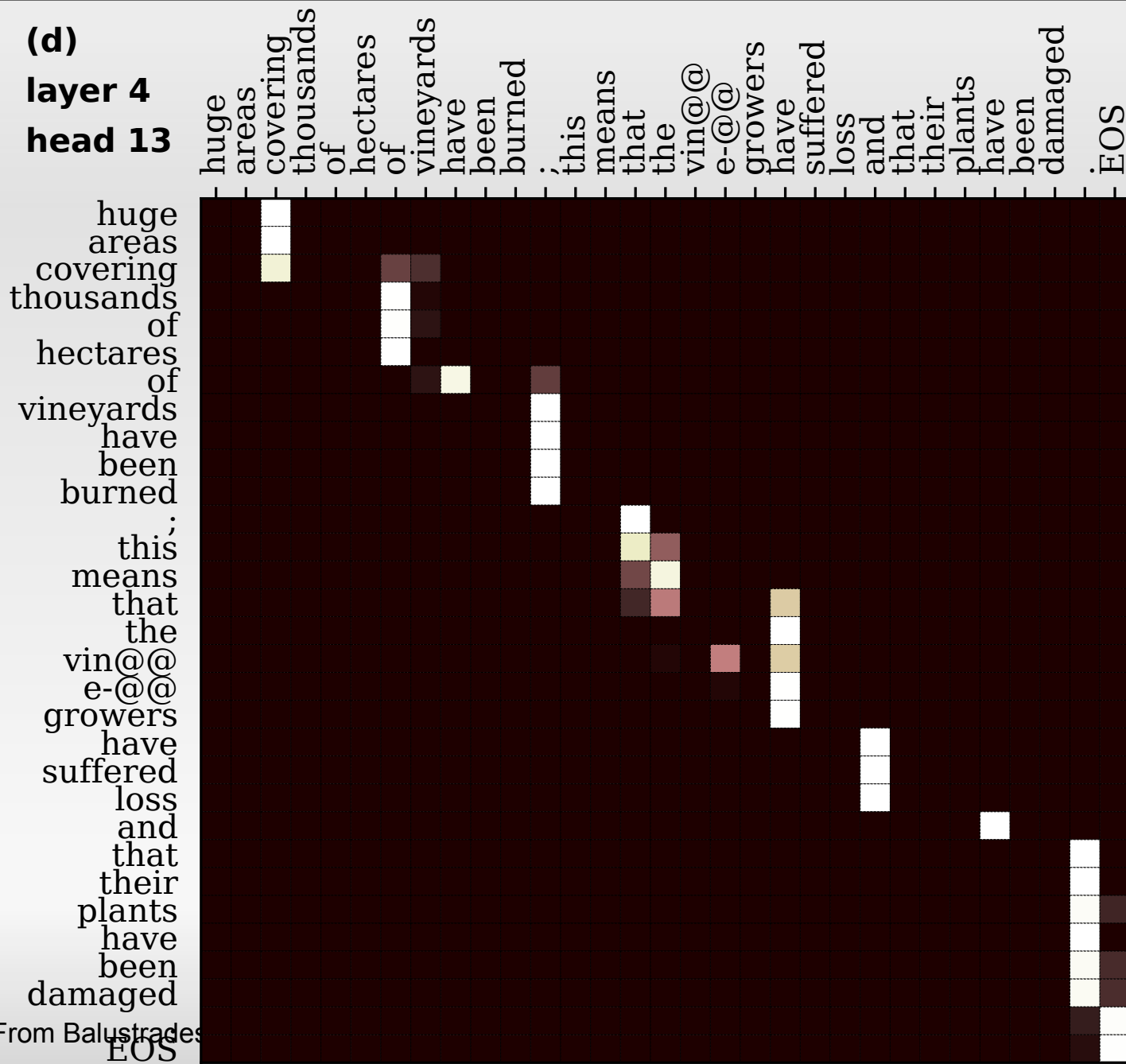


# Shifted diagonal (previous word)

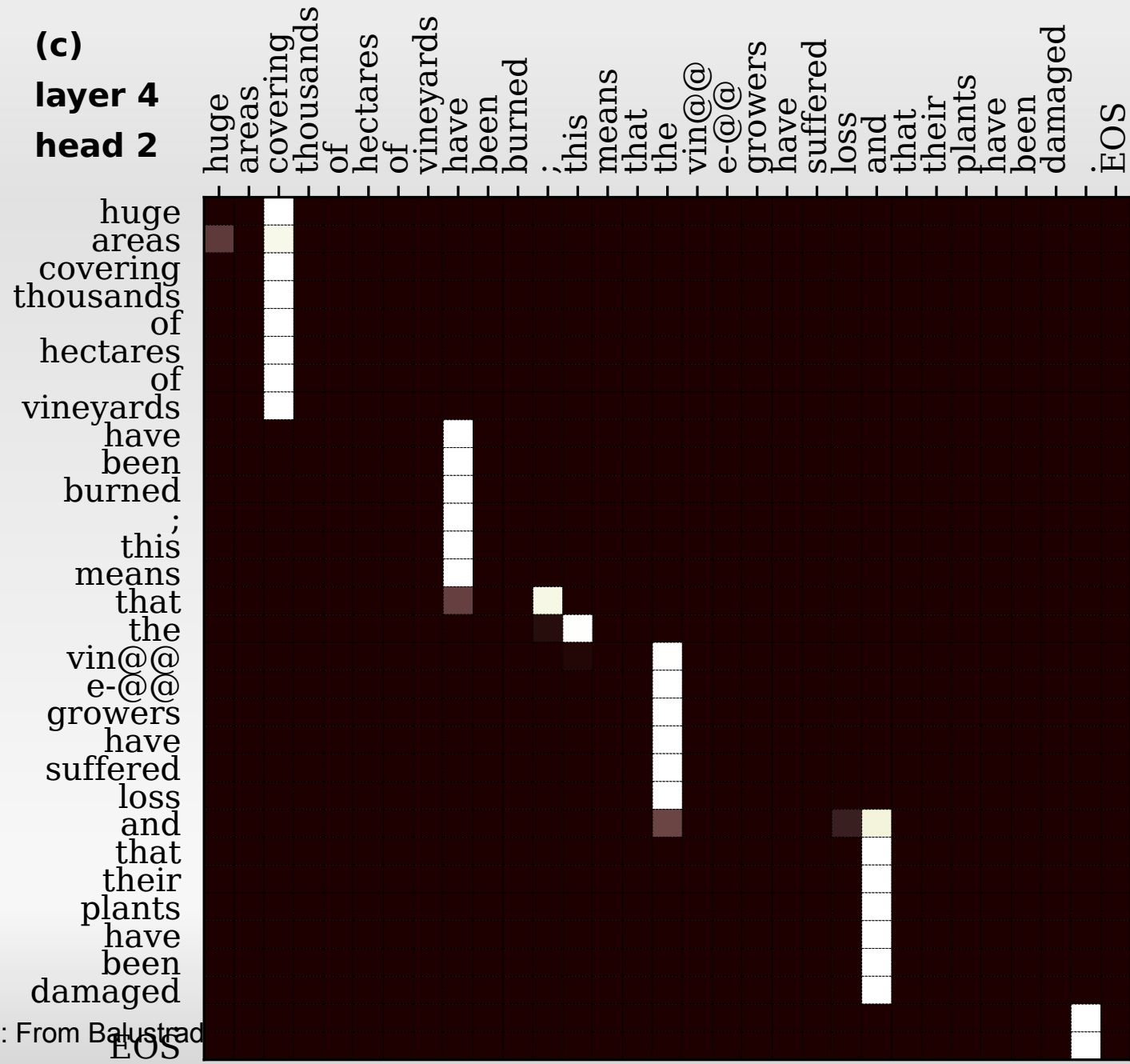




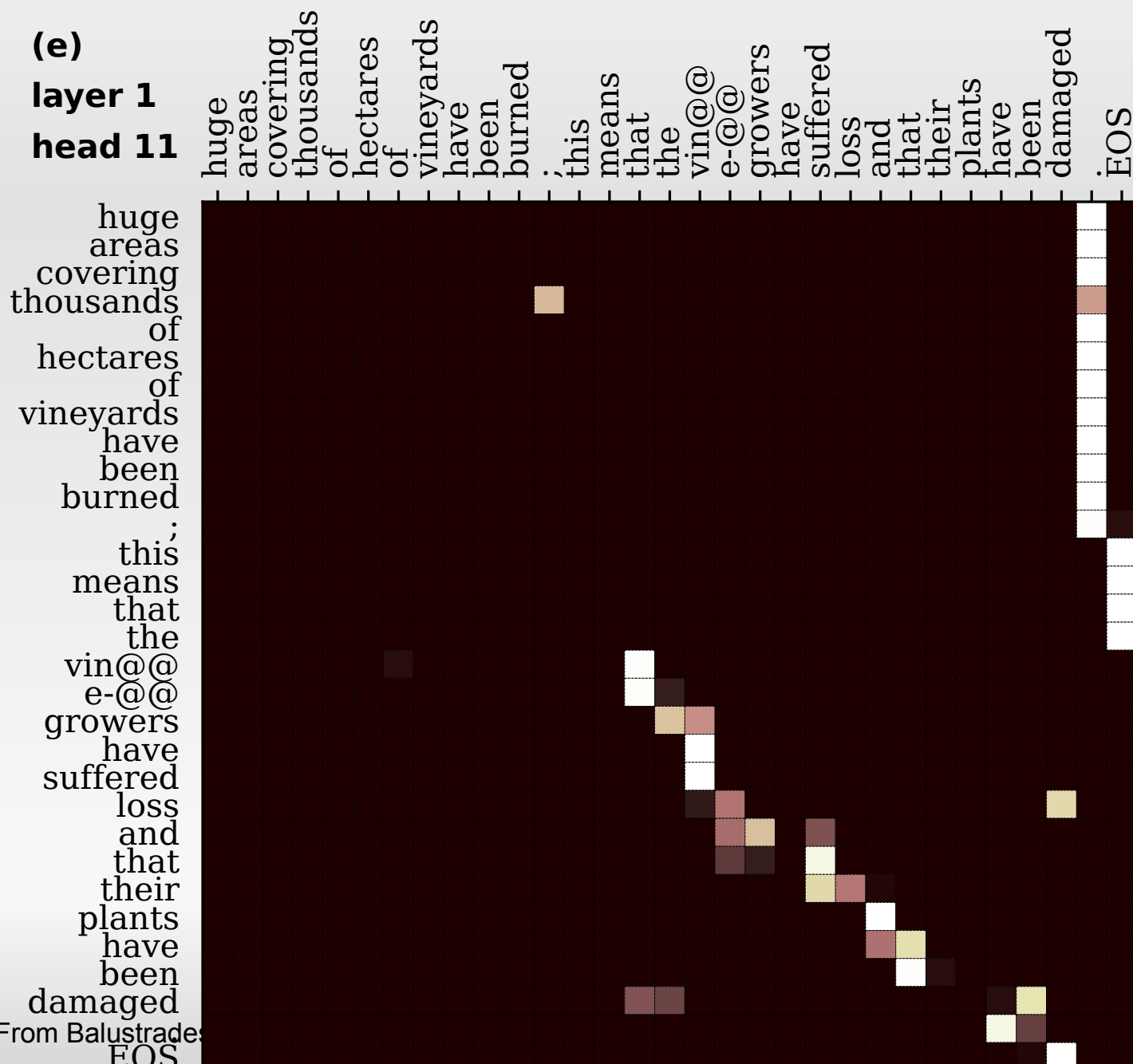
# Short balusters



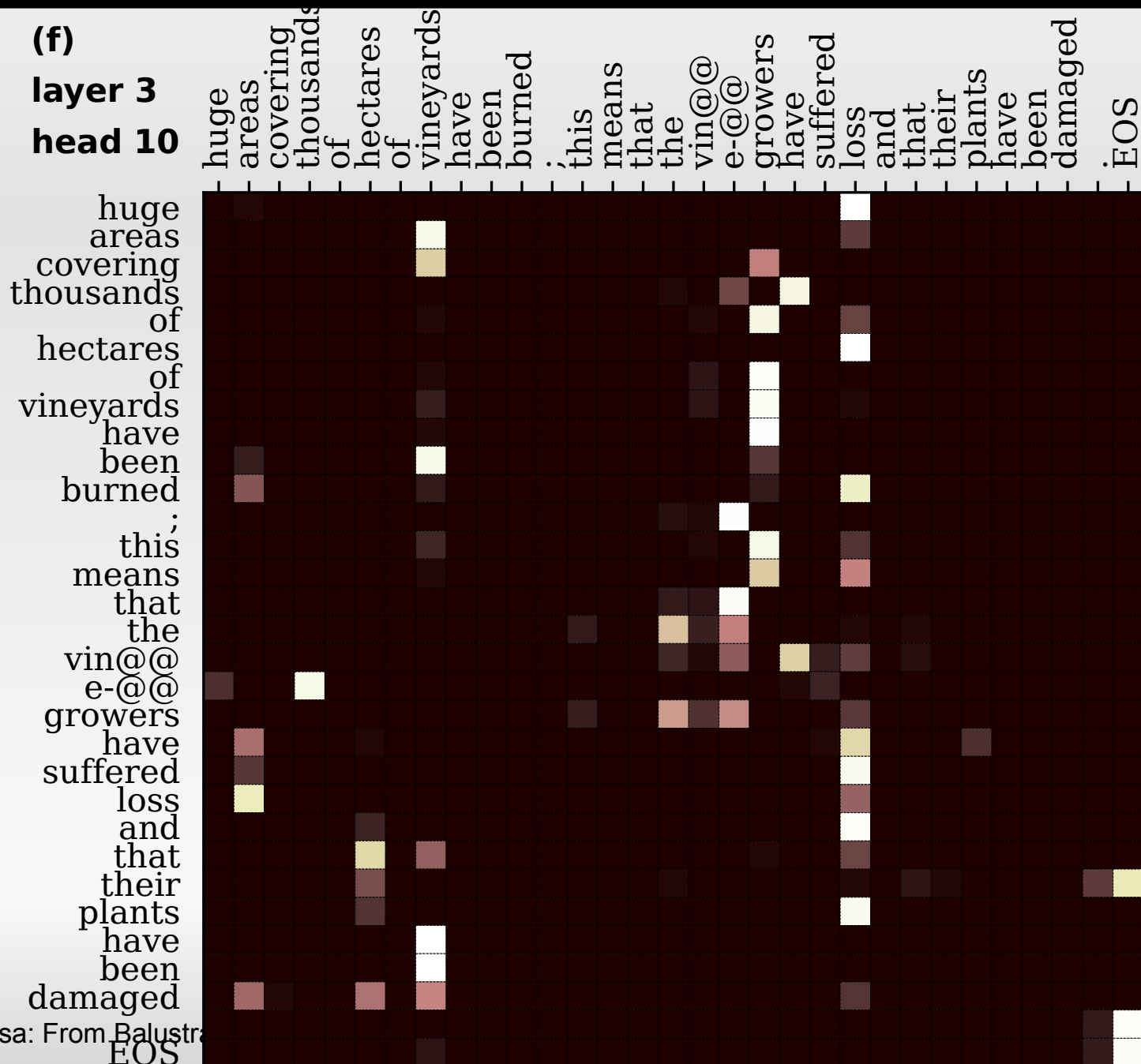
# Long balusters



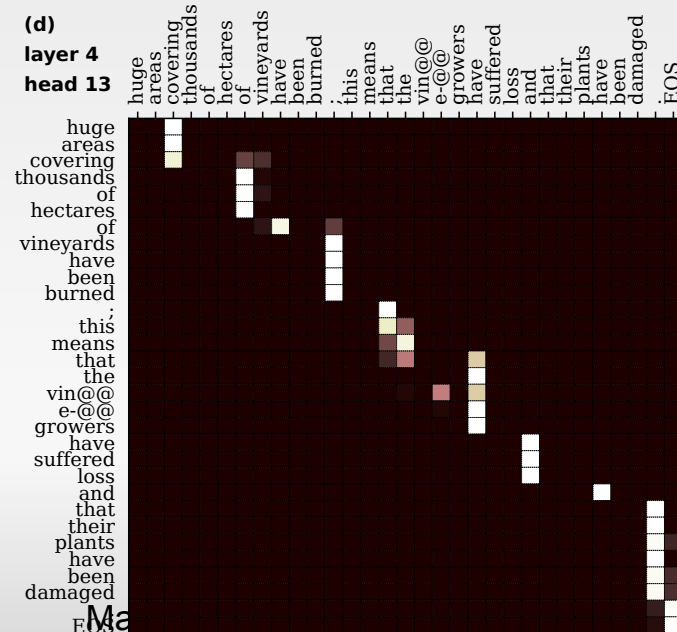
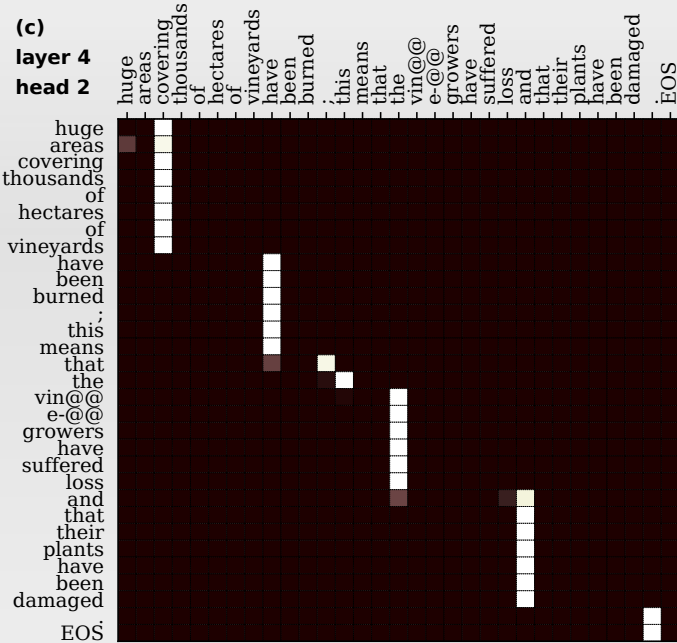
# Partial balustrades + attend to end



# Scattered attention (uninterpreted)



# Phrase candidates & scoring



- keep only max on each line
- phrase candidate
  - each contiguous baluster
  - sequence of words attending to the same position
- phrase score
  - average attention weight
  - sum over all layers and heads
  - short phrases more common  
→ equalization

linguistically uninformed!

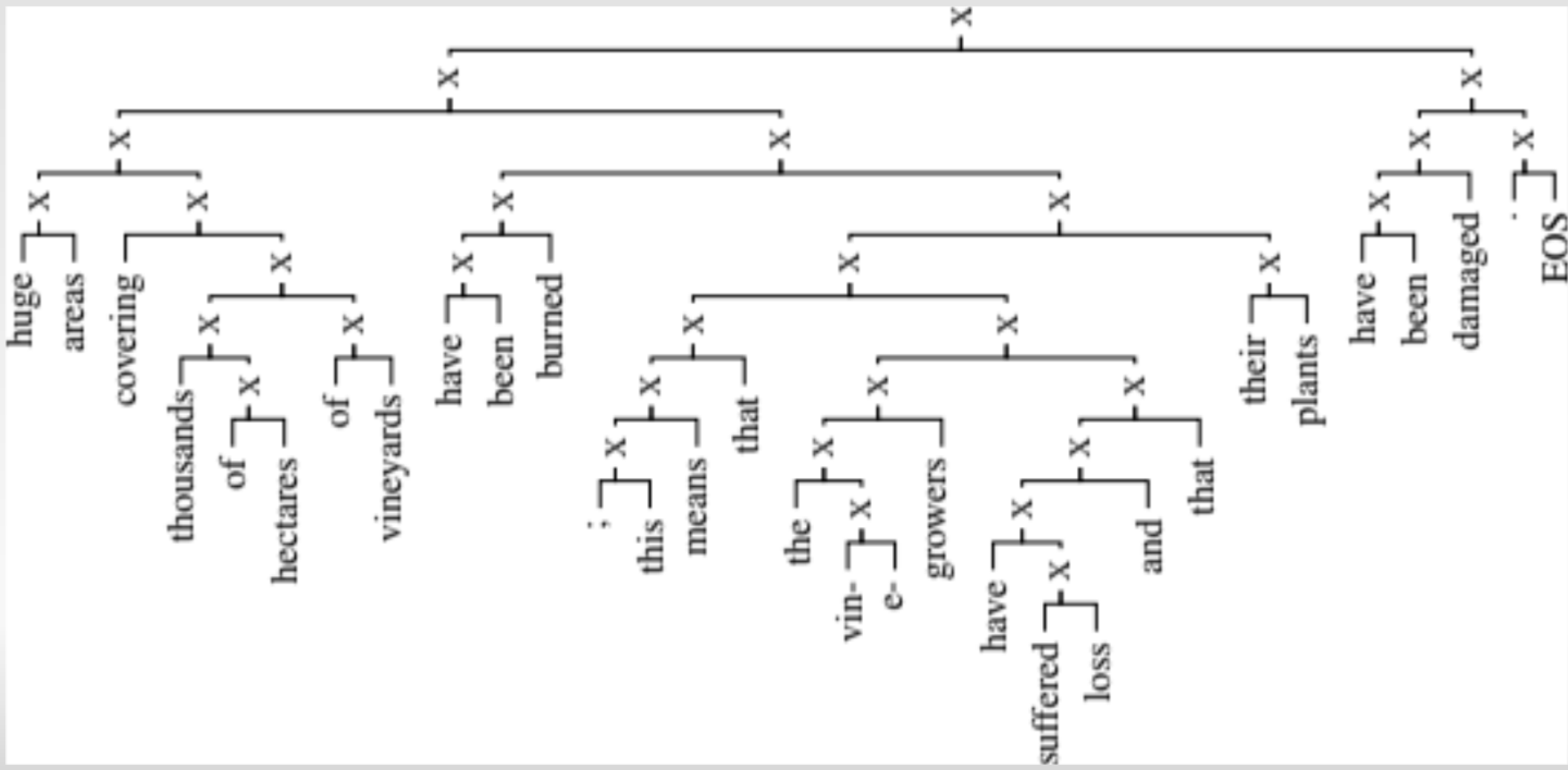
# Binary constituency CKY parsing

- standard recursive algorithm
- constructs a binary constituency tree which maximizes the sum of scores of phrases in the tree
- split each phrase into a pair of subphrases so as to maximize the sum of phrase scores
- linguistically uninformed!

$$S_{a,b} = \max_k \frac{S_{a,k} + S_{k+1,b} + W_{a,k} + W_{k+1,b}}{4}$$

# Results

Huge areas covering thousands of hectares of vineyards have been burned; this means that the vine-growers have suffered loss and that their plants have been damaged.



# Results

English			
system	precision	recall	F1 score
rbal	30.1%	24.3%	26.8%
lbal	27.8%	20.8%	23.8%
rand.init	25.1%	20.0%	22.3%
en → de	35.4%	30.6%	32.8%
en → fr	35.4%	30.2%	32.6%

German			
system	precision	recall	F1 score
rbal	39.1%	31.3%	34.8%
lbal	38.1%	27.6%	32.0%
rand.init	33.7%	25.9%	29.3%
de → en	46.1%	39.6%	42.6%
de → fr	46.7%	40.9%	43.6%

French			
system	precision	recall	F1 score
rbal	34.3%	28.7%	31.3%
lbal	32.5%	25.4%	28.5%
rand.init	26.1%	24.4%	25.3%
fr → en	44.4%	39.7%	41.9%
fr → de	46.9%	41.7%	44.2%

Table 2: Scores of baseline trees and our extracted trees using all attention heads, evaluated against standard syntactic parse trees.



# Summary

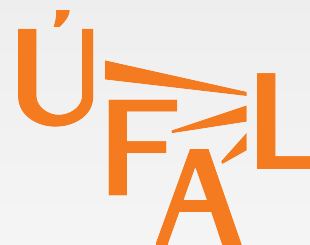
- Transformer NMT encoder self-attentions
  - diagonals, shifted diagonals, scattered attention...
  - balustrades: can be interpreted as phrases
- Linguistically uninformed syntax extraction
  - baluster → phrase, attention weight → phrase score
  - binary constituency parsing using CKY
  - no training, no hyperparameters, using all heads
    - see the paper for subselecting only some heads
- Resulting structures are quite syntactically sane
  - F1 score 6 – 13 points above baseline (30% → 40%)

# Thank you for your attention

David Mareček, Rudolf Rosa  
marecek@ufal.mff.cuni.cz, rosa@ufal.mff.cuni.cz

## **From Balustrades to Pierre Vinken: Looking for Syntax in Transformer Self-Attentions**

Charles University, Prague  
Faculty of Mathematics and Physics  
Institute of Formal and Applied Linguistics



[ufal.cz/grants/lzd](https://ufal.cz/grants/lzd)