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

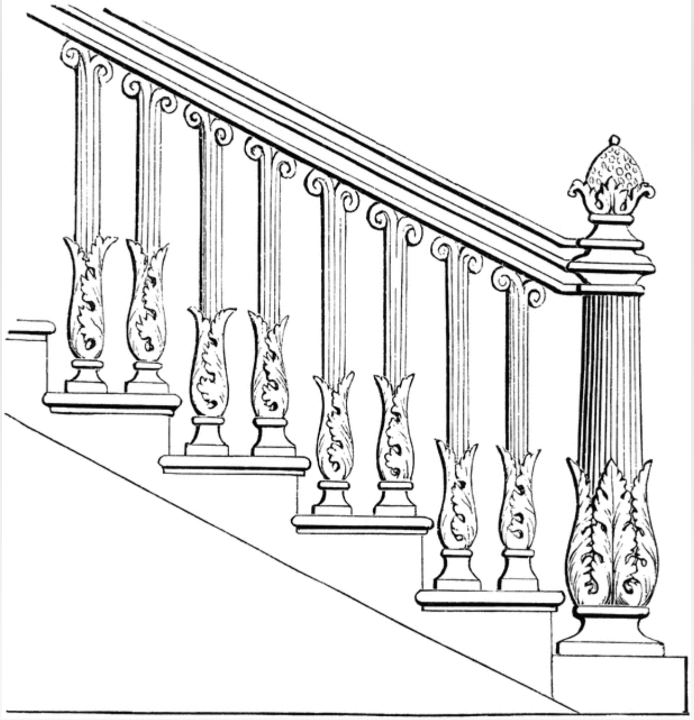
From Balustrades to Pierre Vinken:

Looking for Syntax in Transformer Self-Attentions

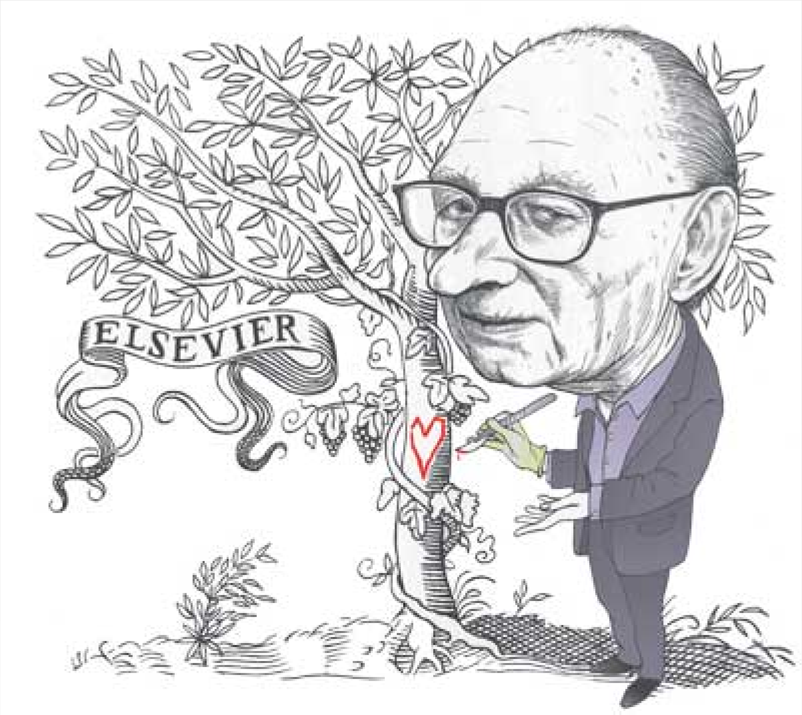
Charles University, Prague
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

BlackboxNLP Workshop, Firenze, 1 August 2019

From balustrades to Pierre Vinken

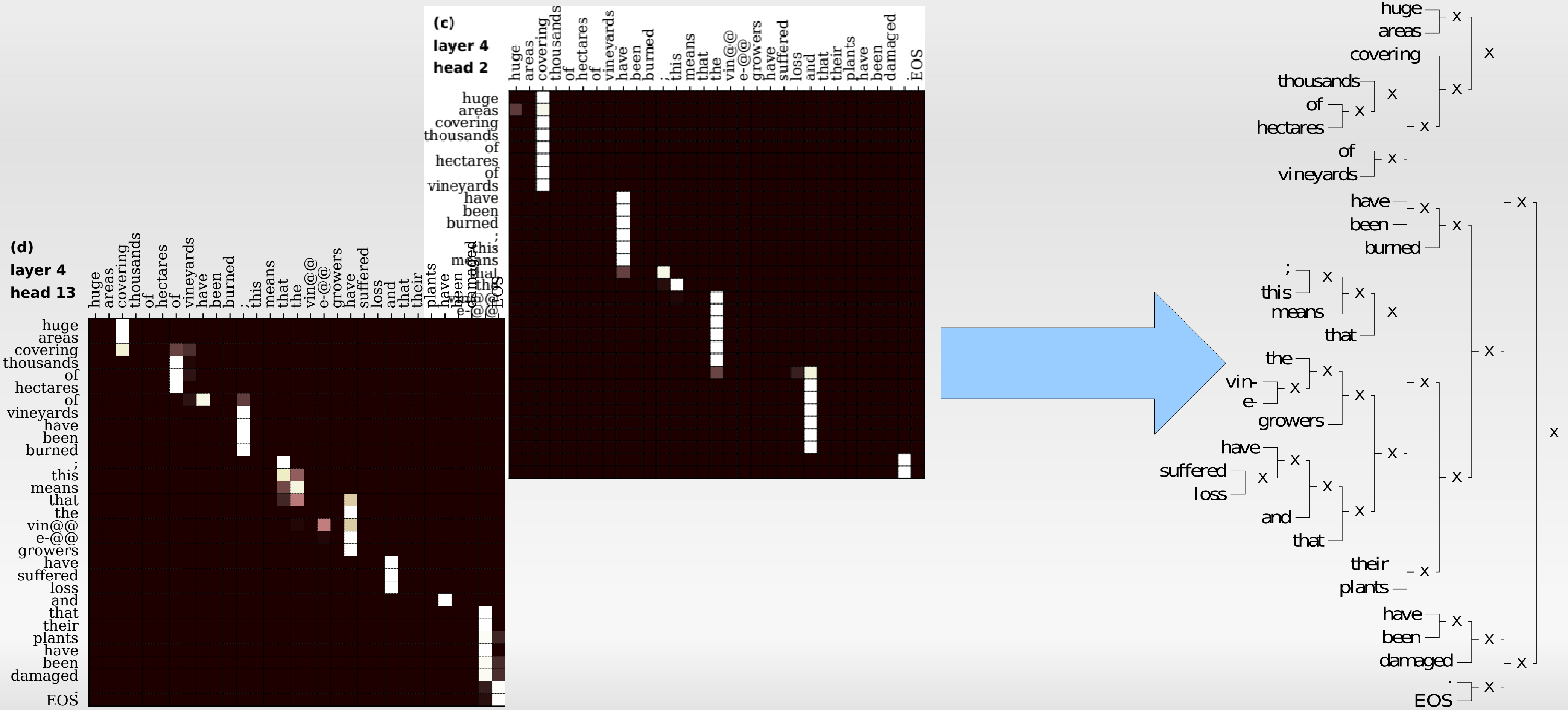


http://clipart-library.com/clipart/28144.htm



by Jan Hein van Dierendonck

Transformer self-attentions → syntactic trees



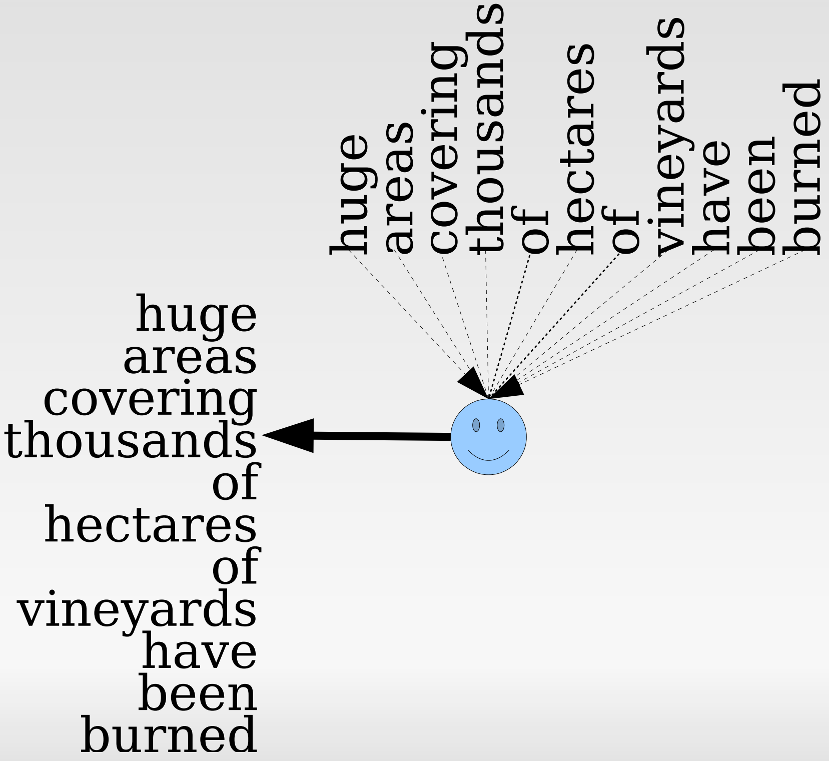
Observation

Observation

- Common pattern in Transformer NMT self-attention heads

Observation

- Common pattern in Transformer NMT self-attention heads

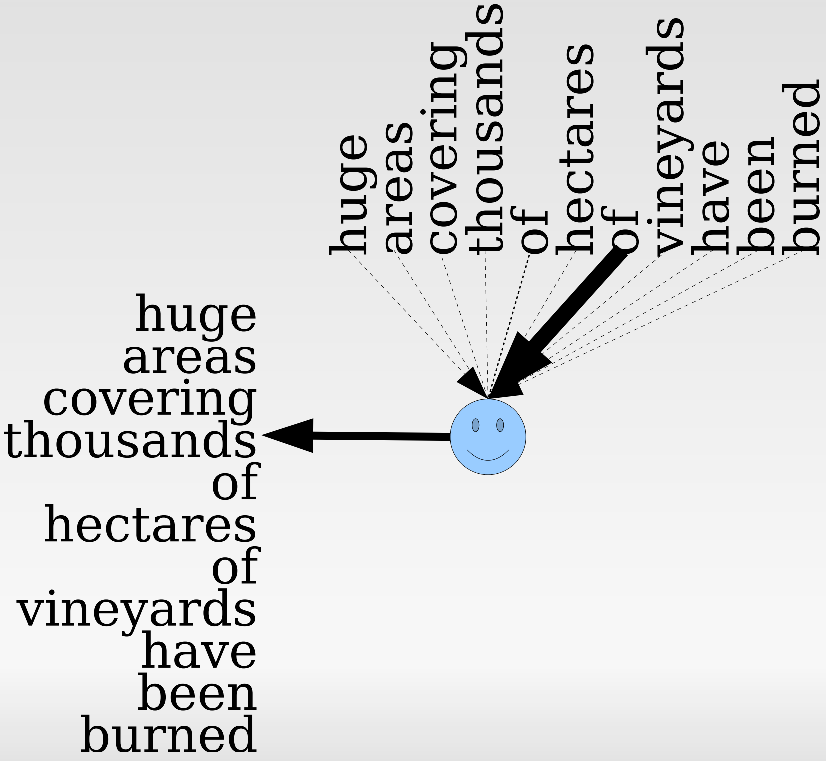


huge areas covering thousands of hectares of vineyards have been burned

huge
areas
covering
thousands
of
hectares
of
vineyards
have
been
burned

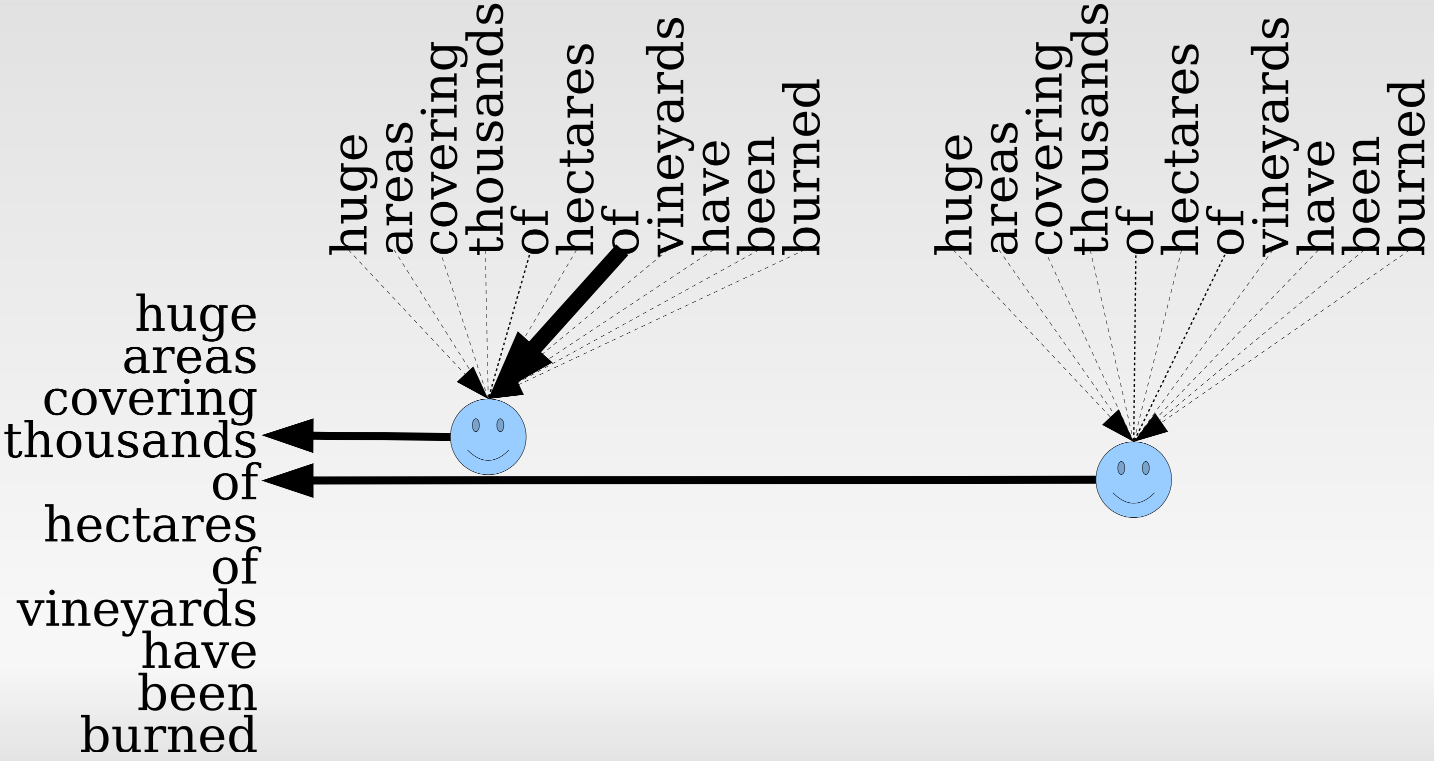
Observation

- Common pattern in Transformer NMT self-attention heads



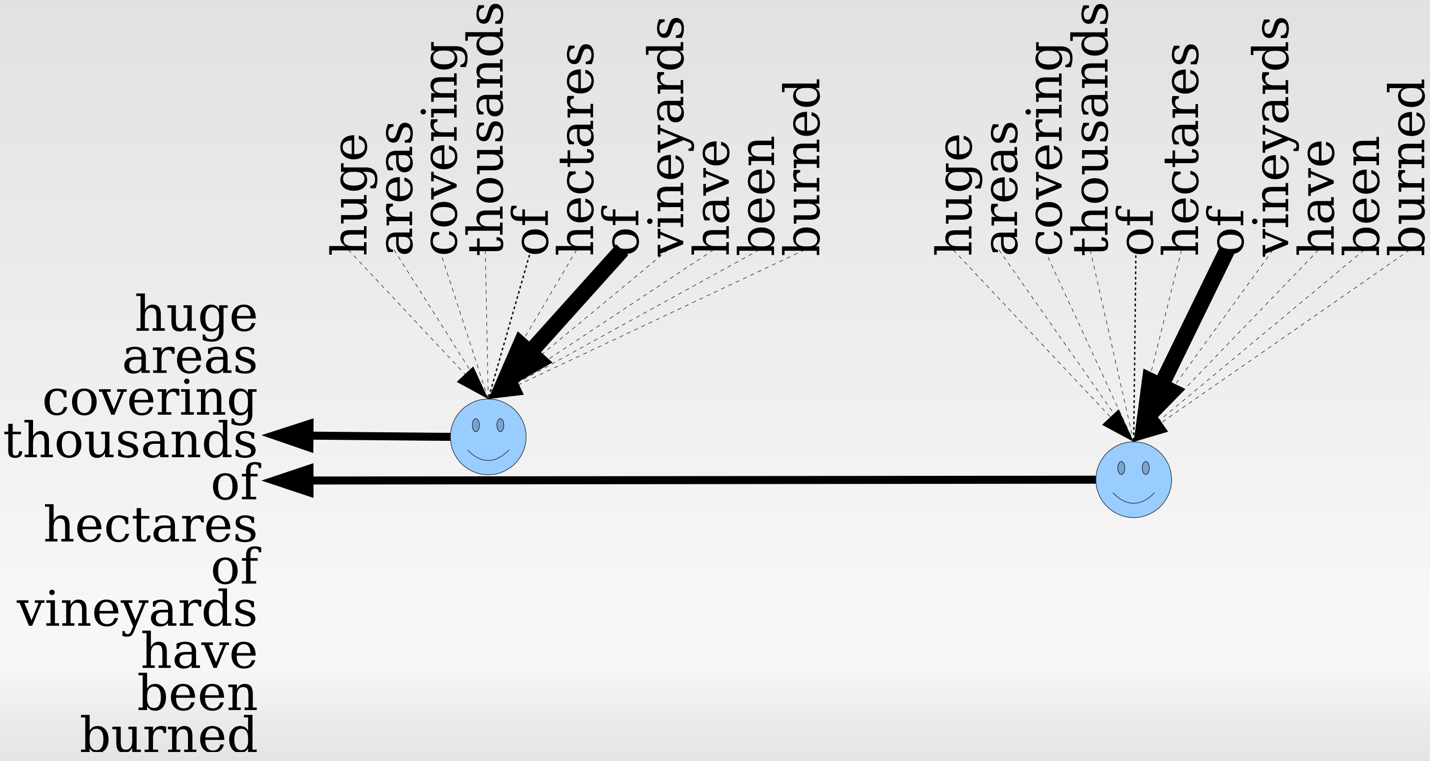
Observation

- Common pattern in Transformer NMT self-attention heads



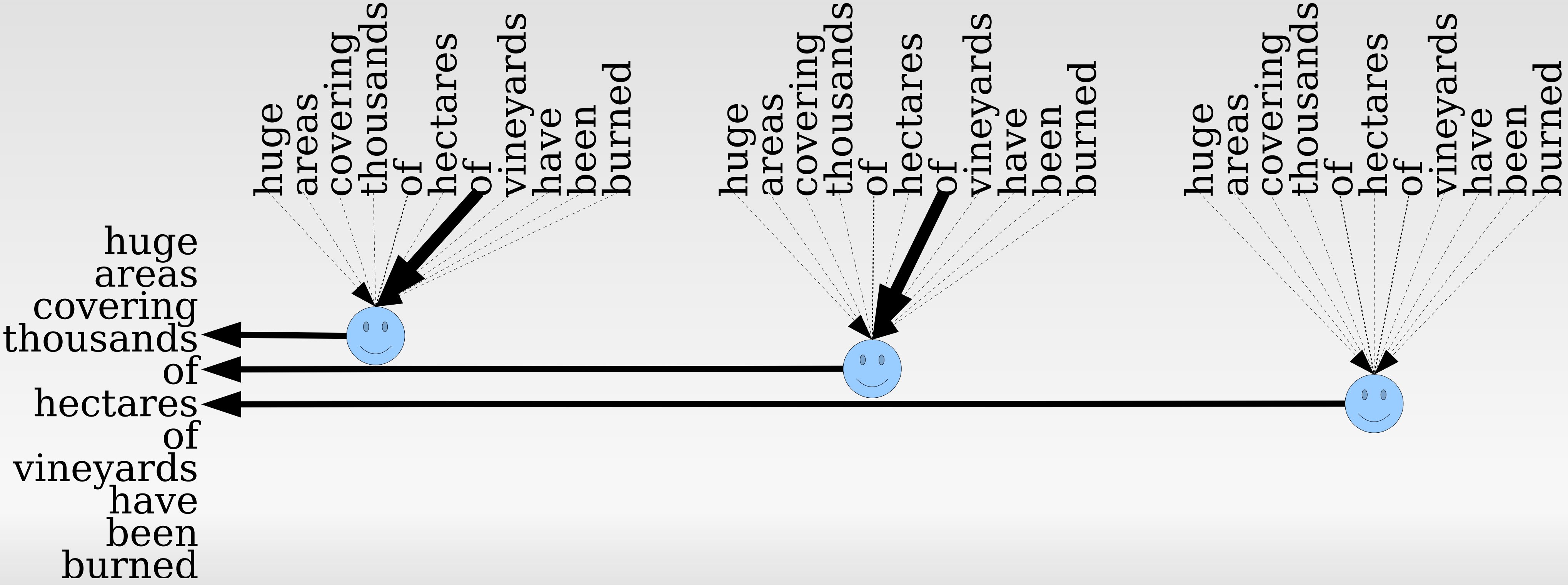
Observation

- Common pattern in Transformer NMT self-attention heads



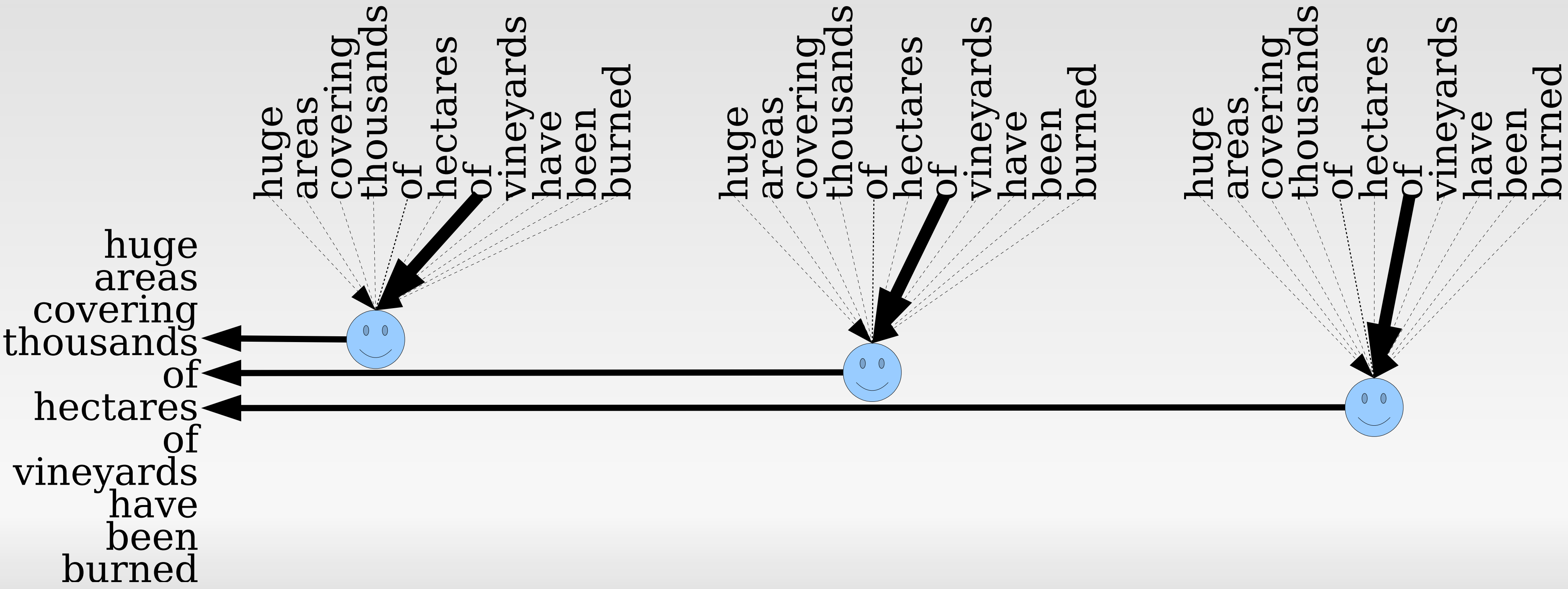
Observation

- Common pattern in Transformer NMT self-attention heads



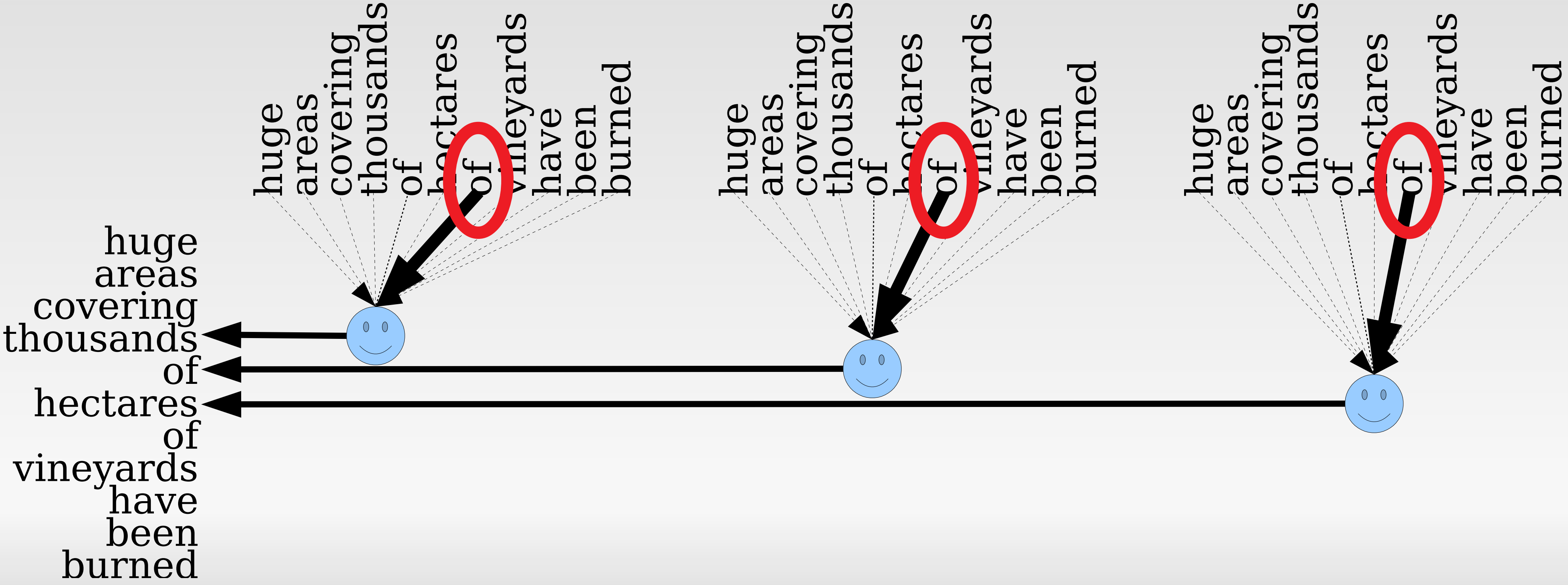
Observation

- Common pattern in Transformer NMT self-attention heads



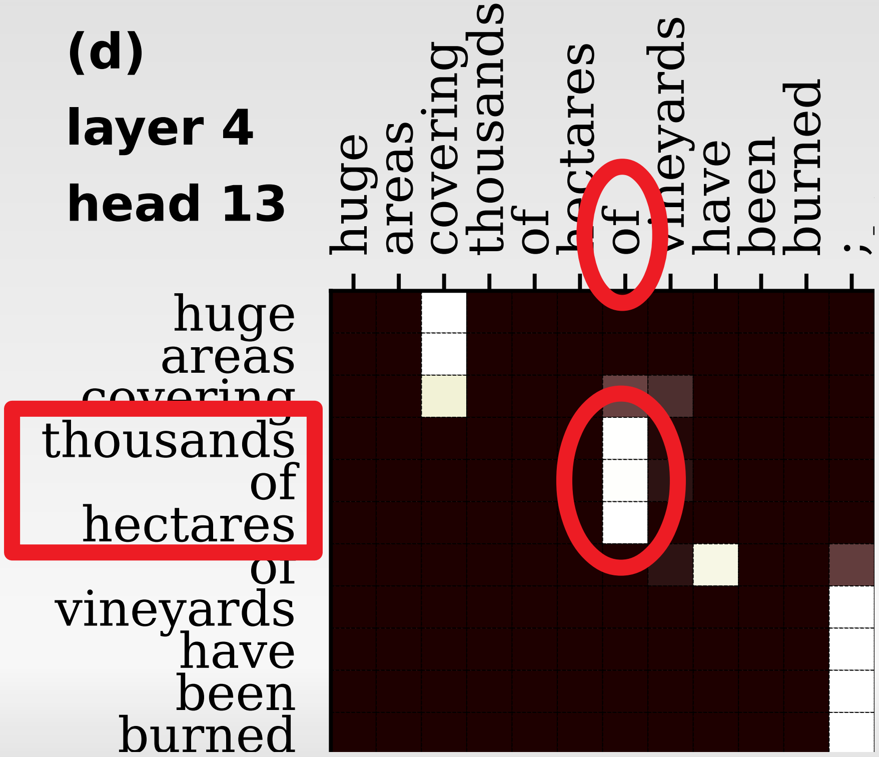
Observation

- Common pattern in Transformer NMT self-attention heads



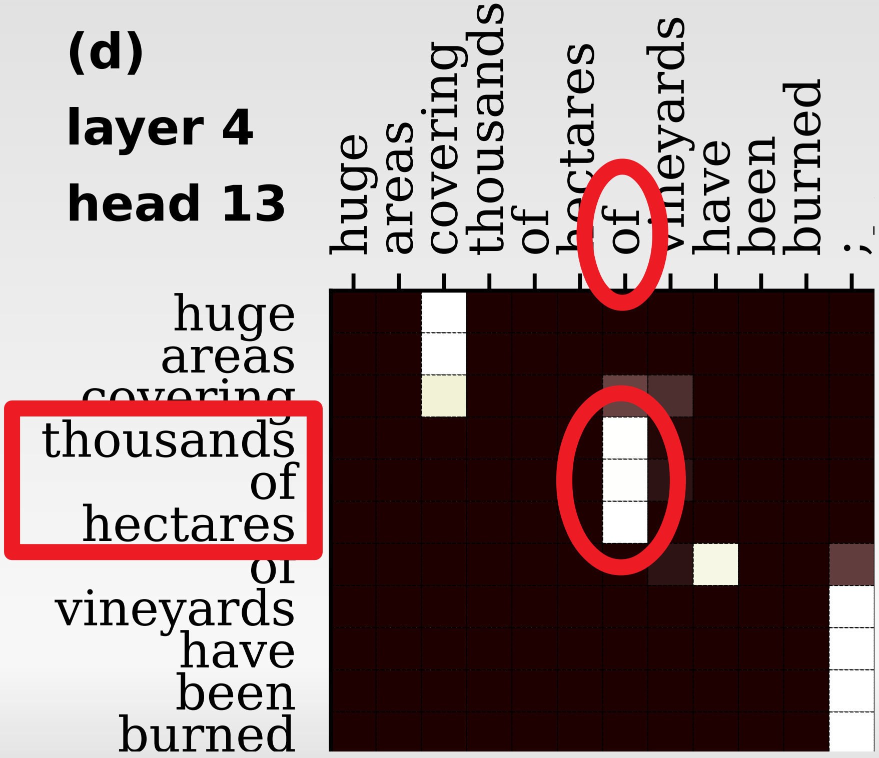
Observation

- Common pattern in Transformer NMT self-attention heads



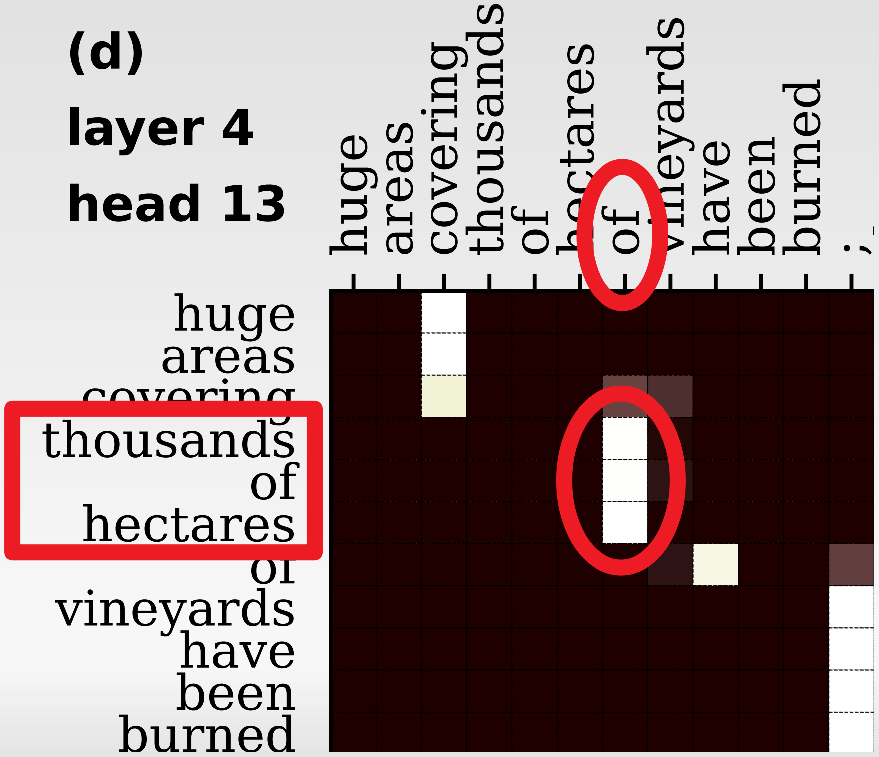
Observation

- Common pattern in Transformer NMT self-attention heads
  - “balusters”



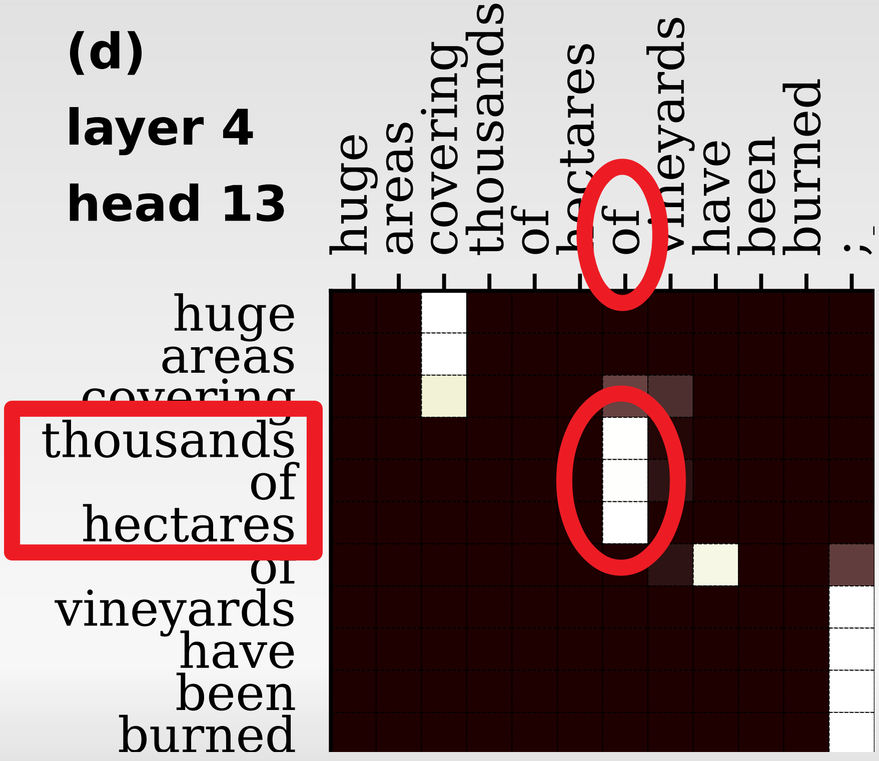
Observation

- Common pattern in Transformer NMT self-attention heads
    - “balusters”
- Resemble syntactic phrases



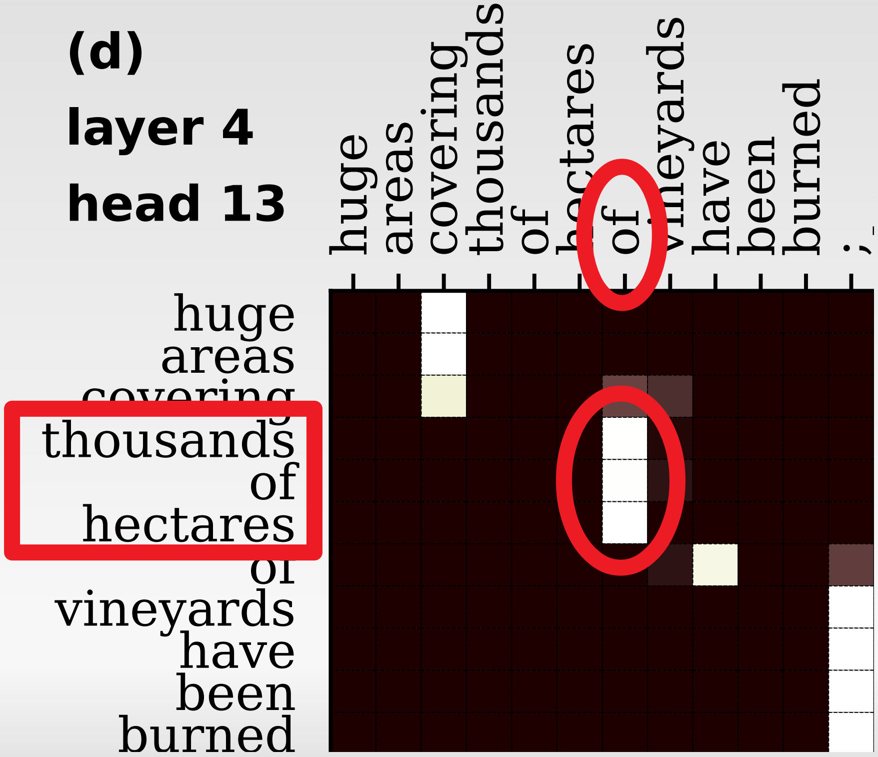
Observation

- Common pattern in Transformer NMT self-attention heads
  - “balusters”
- Resemble syntactic phrases



Observation

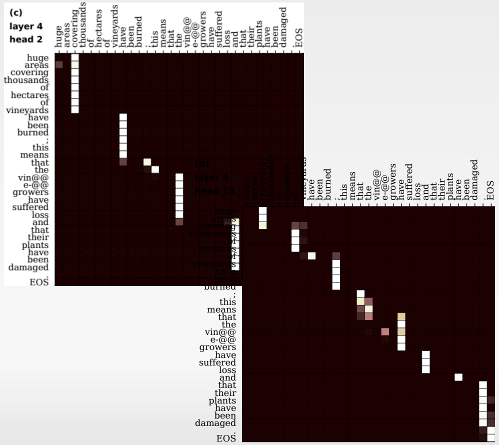
- Common pattern in Transformer NMT self-attention heads
  - “balusters”
- Resemble syntactic phrases
  - To what extent?
    → That’s our research question!



Approach

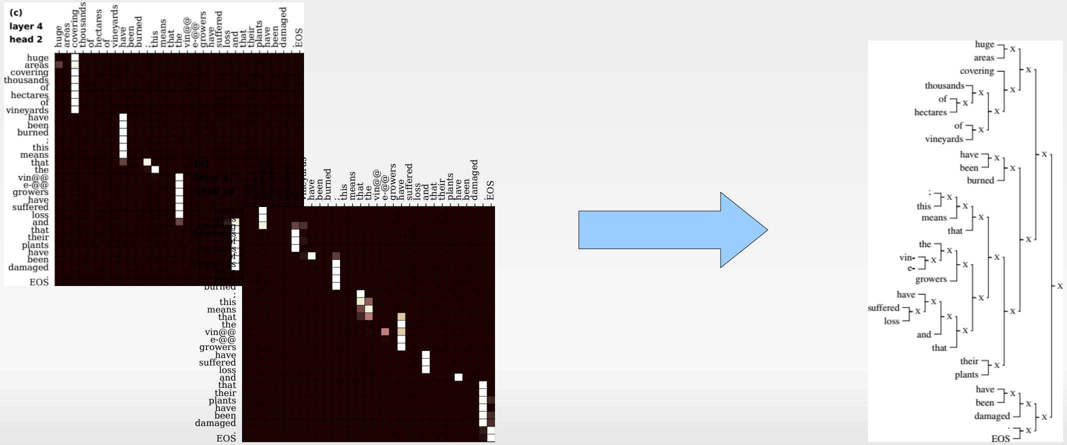
Approach

1. Balusters → phrase candidates



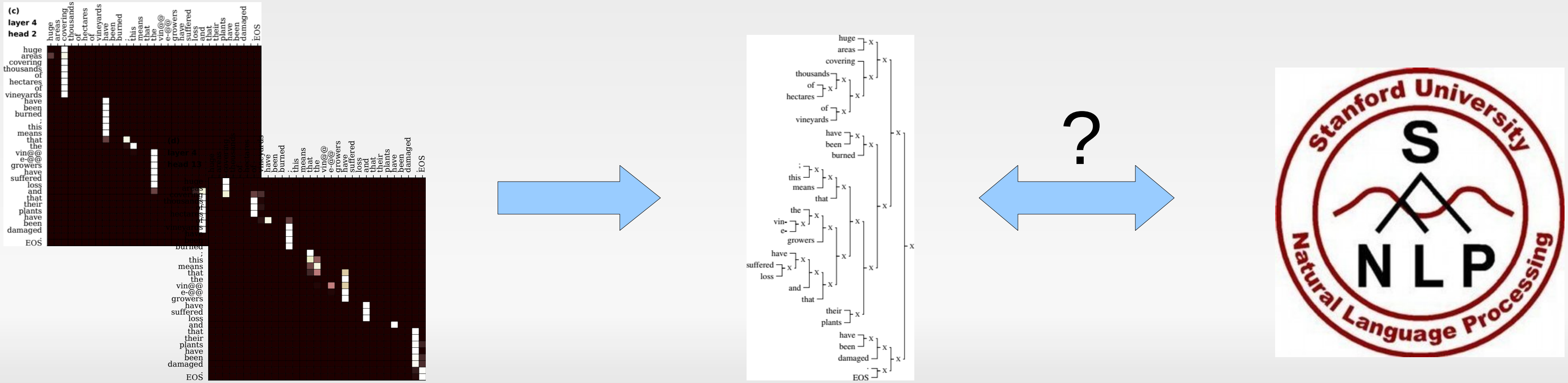
Approach

1. Balusters → phrase candidates
2. Phrase candidates → constituency tree
   - Linguistically uninformed algorithm


(c)
layer 4
head 2
huge areas covering thousands of hectares of vineyards have been burned ; this means that the vin@@ e-@@ growers have suffered loss and that their plants have been damaged . EOS
huge
areas
covering
thousands
of
hectares
of
vineyards
have
been
burned
;
this
means
that
the
vin-
e-
growers
have
suffered
loss
and
that
their
plants
have
been
damaged
.
EOS
x

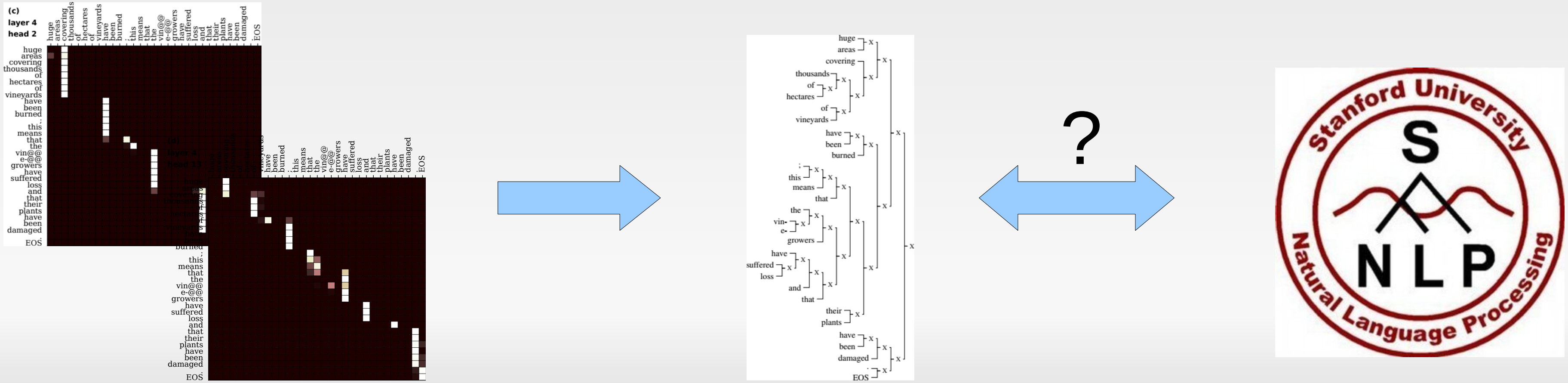
Approach

1. Balusters → phrase candidates
2. Phrase candidates → constituency tree
   - Linguistically uninformed algorithm
3. Compare to standard syntactic trees



Approach

1. Balusters → phrase candidates
2. Phrase candidates → constituency tree
   - Linguistically uninformed algorithm
3. Compare to standard syntactic trees: ~40%; baseline ~30%



Experiment setup

- Balusters: Transformer NMT system
  - Encoder: 6 layers x 16 heads

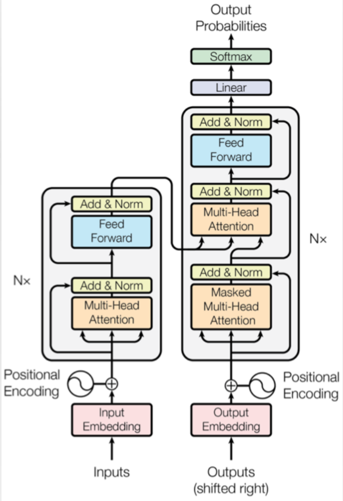


Figure 1: The Transformer - model architecture.

Experiment setup

- Balusters: Transformer NMT system
  - Encoder: 6 layers x 16 heads
  - Europarl: French ↔ English, German ↔ English, French ↔ German

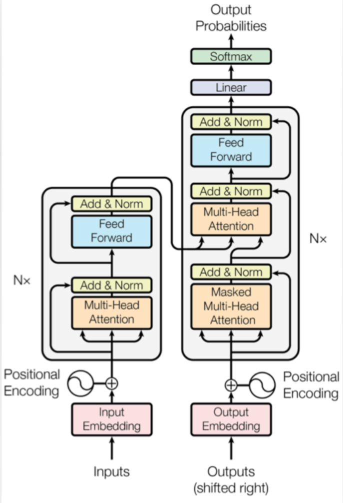


Figure 1: The Transformer - model architecture.

Experiment setup

- Balusters: Transformer NMT system
  - Encoder: 6 layers x 16 heads
  - Europarl: French ↔ English, German ↔ English, French ↔ German

European Parliament

- Standard syntactic trees: Stanford parser
  - Penn Treebank, French Treebank, Negra Corpus
  - Only for evaluation

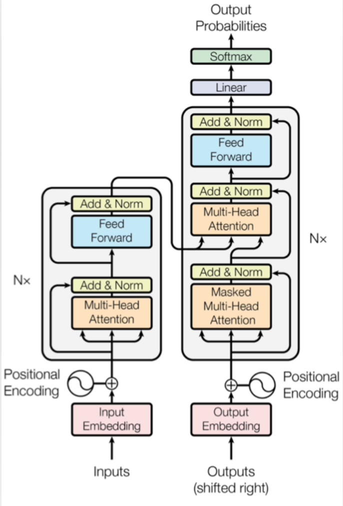
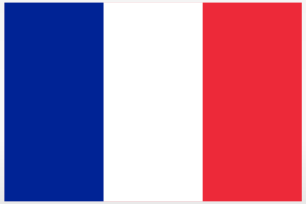
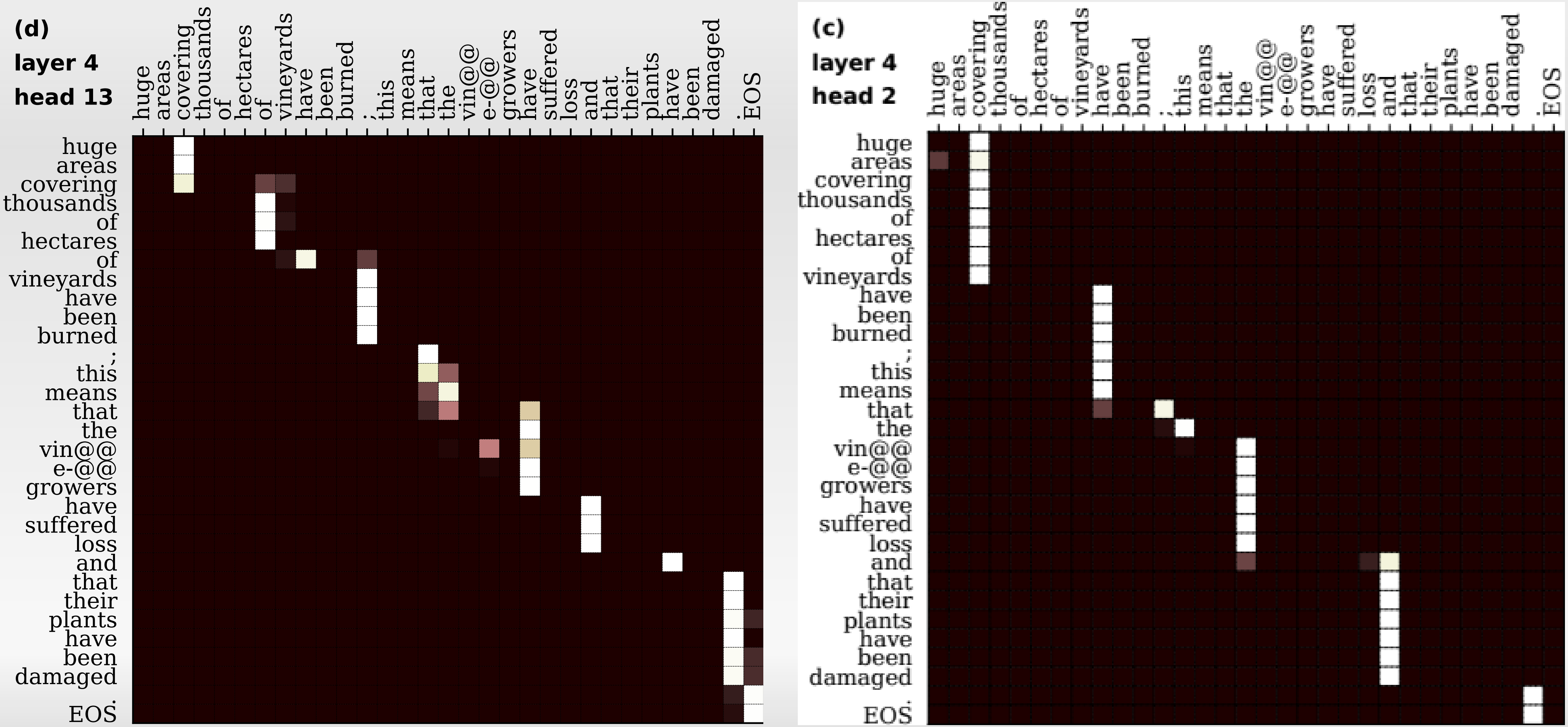


Figure 1: The Transformer - model architecture.

Stanford University S NLP Natural Language Processing

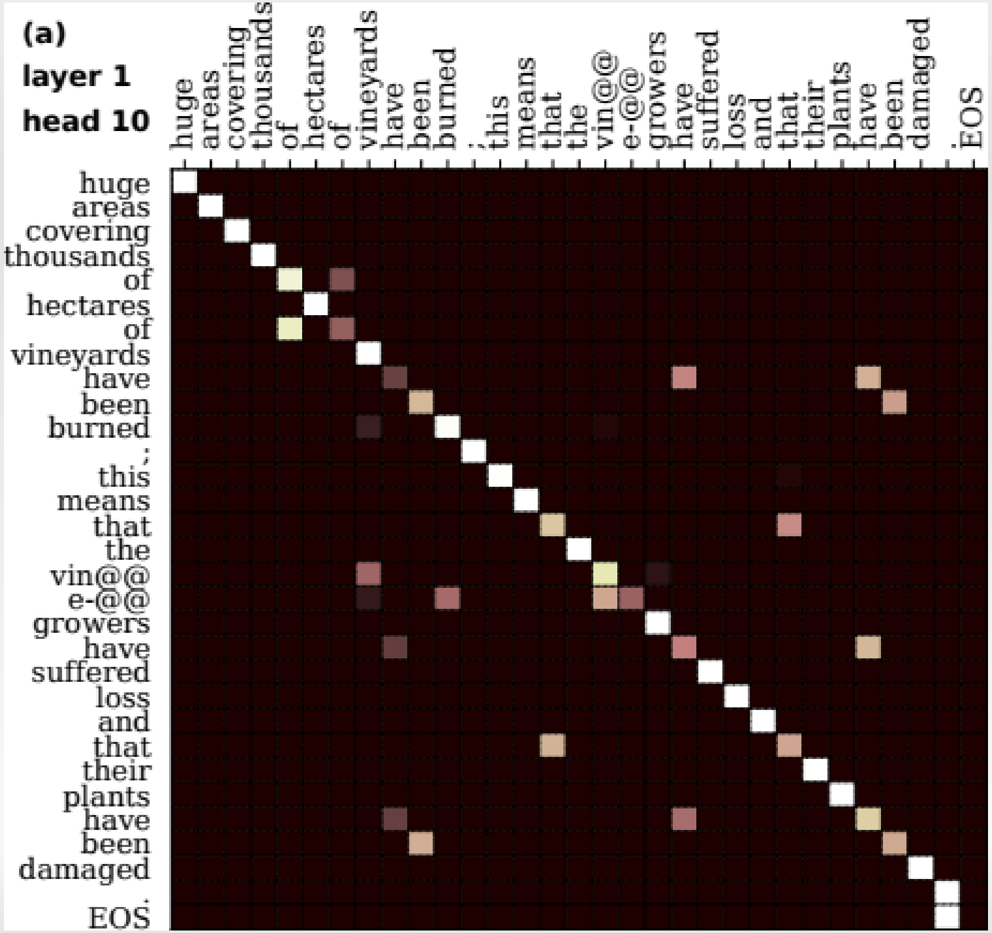


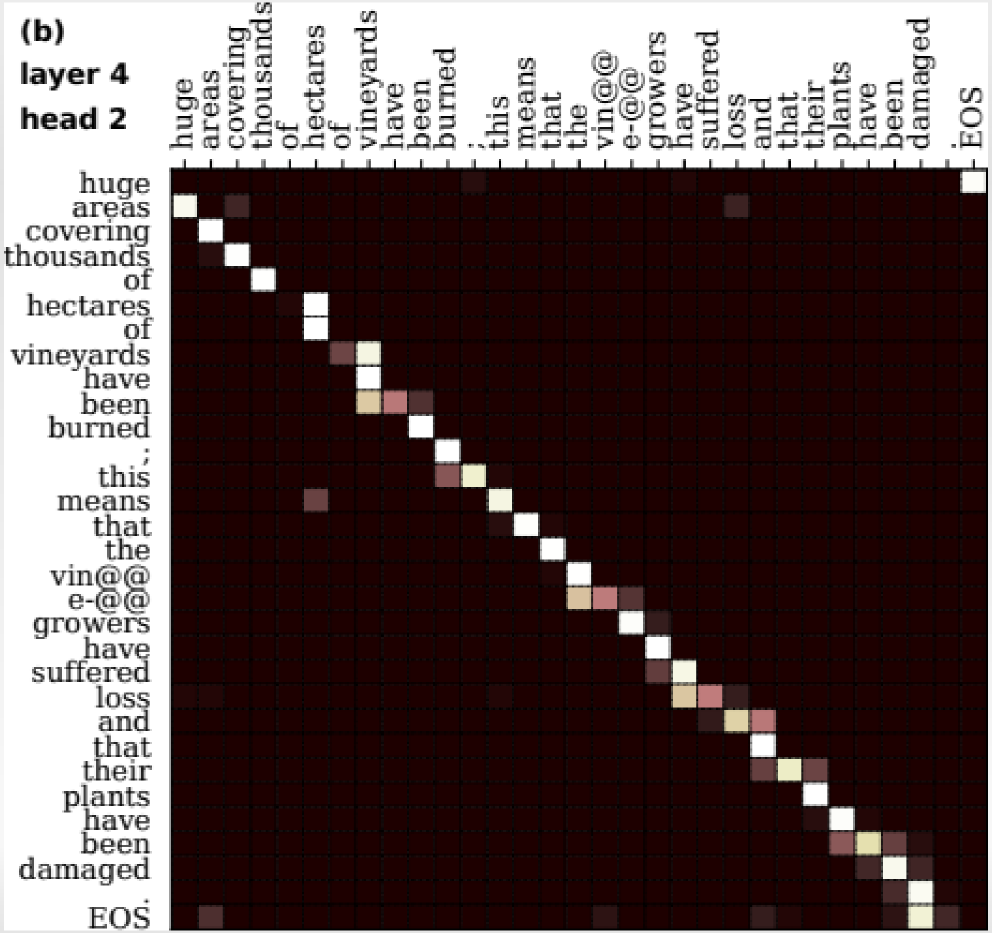
Balustrades (~70% of the attention heads)



Mareček & Rosa: From Balustrades to Pierre Vinken: Looking for Syntax in Transformer Self-Attentions

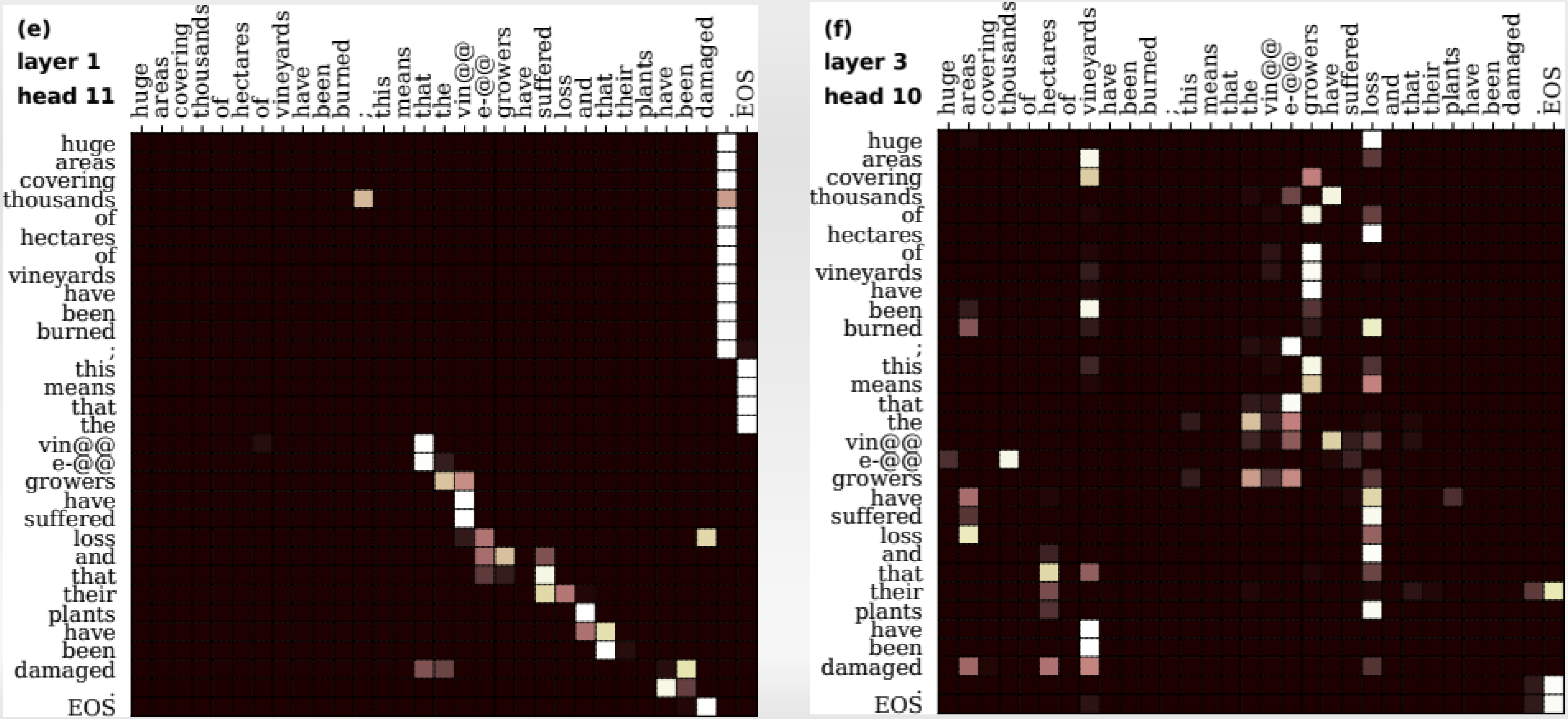
Diagonals (especially 1st layer)





Mareček & Rosa: From Balustrades to Pierre Vinken: Looking for Syntax in Transformer Self-Attentions

Attend to end, mixed, scattered…



Phrase candidates

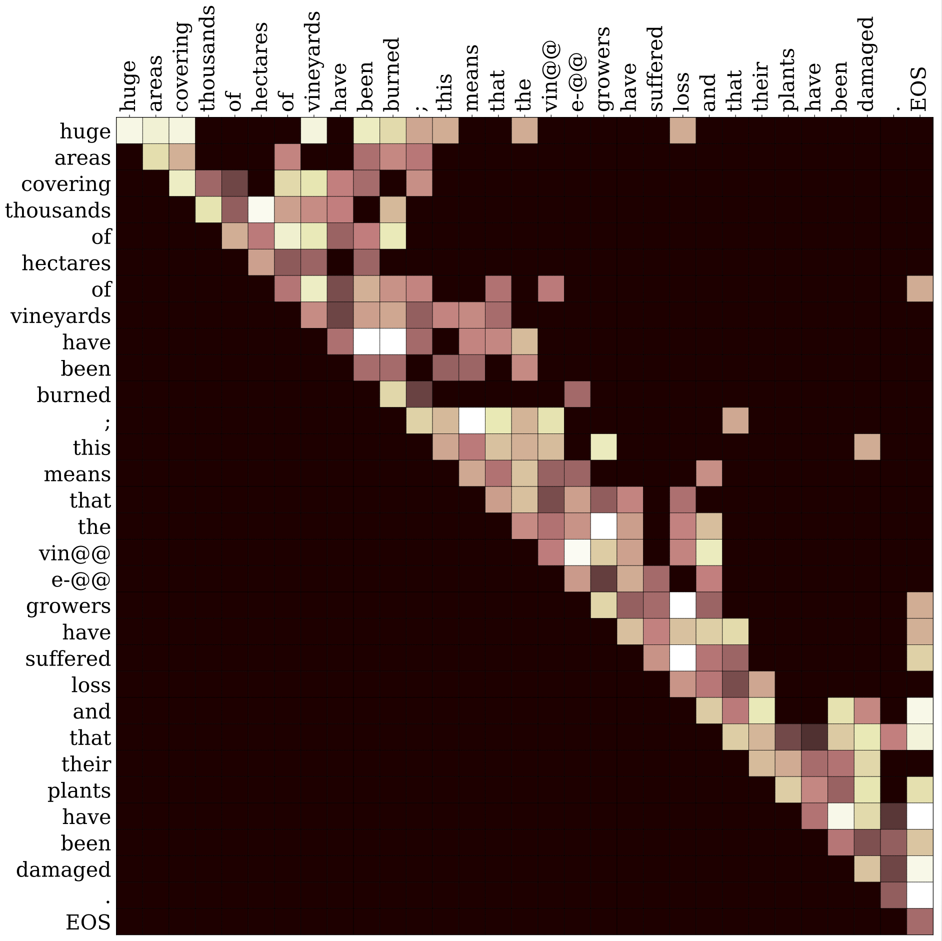
- All balusters of length ≥ 2 from **all** heads
  - Subselecting only some of the heads → see the paper!

Phrase candidates

- All balusters of length ≥ 2 from **all** heads
  - Subselecting only some of the heads → see the paper!
- Phrase score
  - Average attention weight
  - Sum over all heads
  - Equalize over different phrase lengths

Phrase candidates

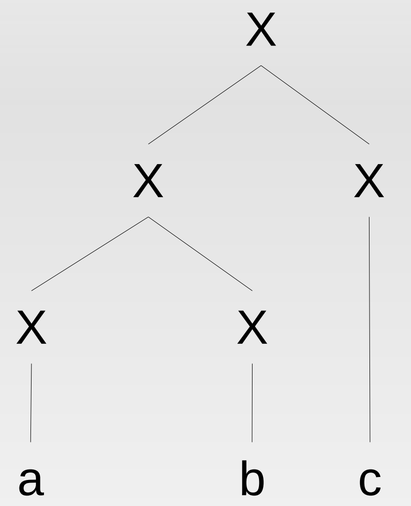
- All balusters of length ≥ 2 from **all** heads
  - Subselecting only some of the heads → see the paper!
- Phrase score
  - Average attention weight
  - Sum over all heads
  - Equalize over different phrase lengths



Phrase candidates → constituency tree

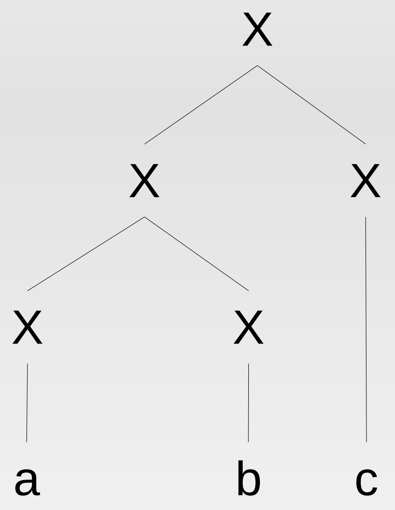
Phrase candidates → constituency tree

- Binary constituency tree



Phrase candidates → constituency tree

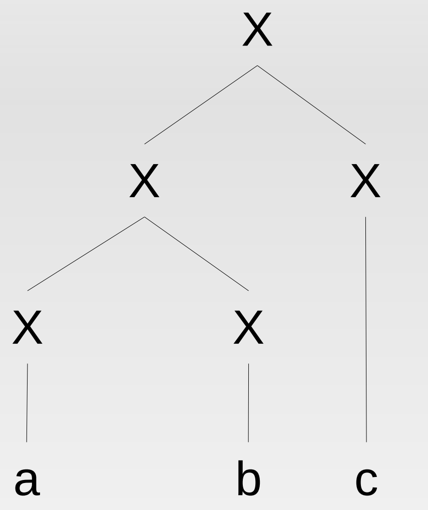
- Binary constituency tree
- Tree score = sum of phrase scores



s(T) = s(ab) + s(abc)

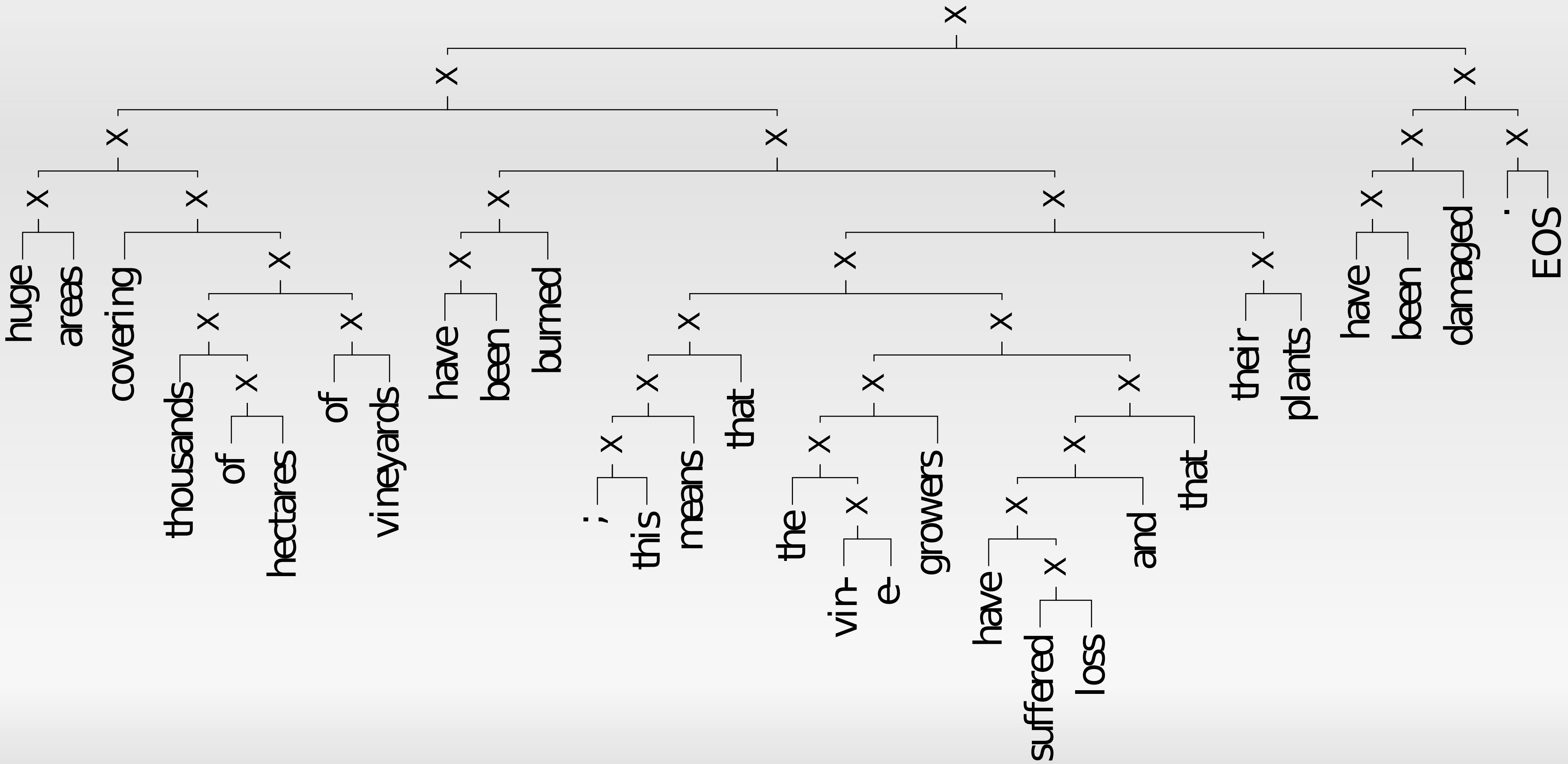
Phrase candidates → constituency tree

- Binary constituency tree
- Tree score = sum of phrase scores
- CKY algorithm
    - Finds tree (set of phrases) with maximal score

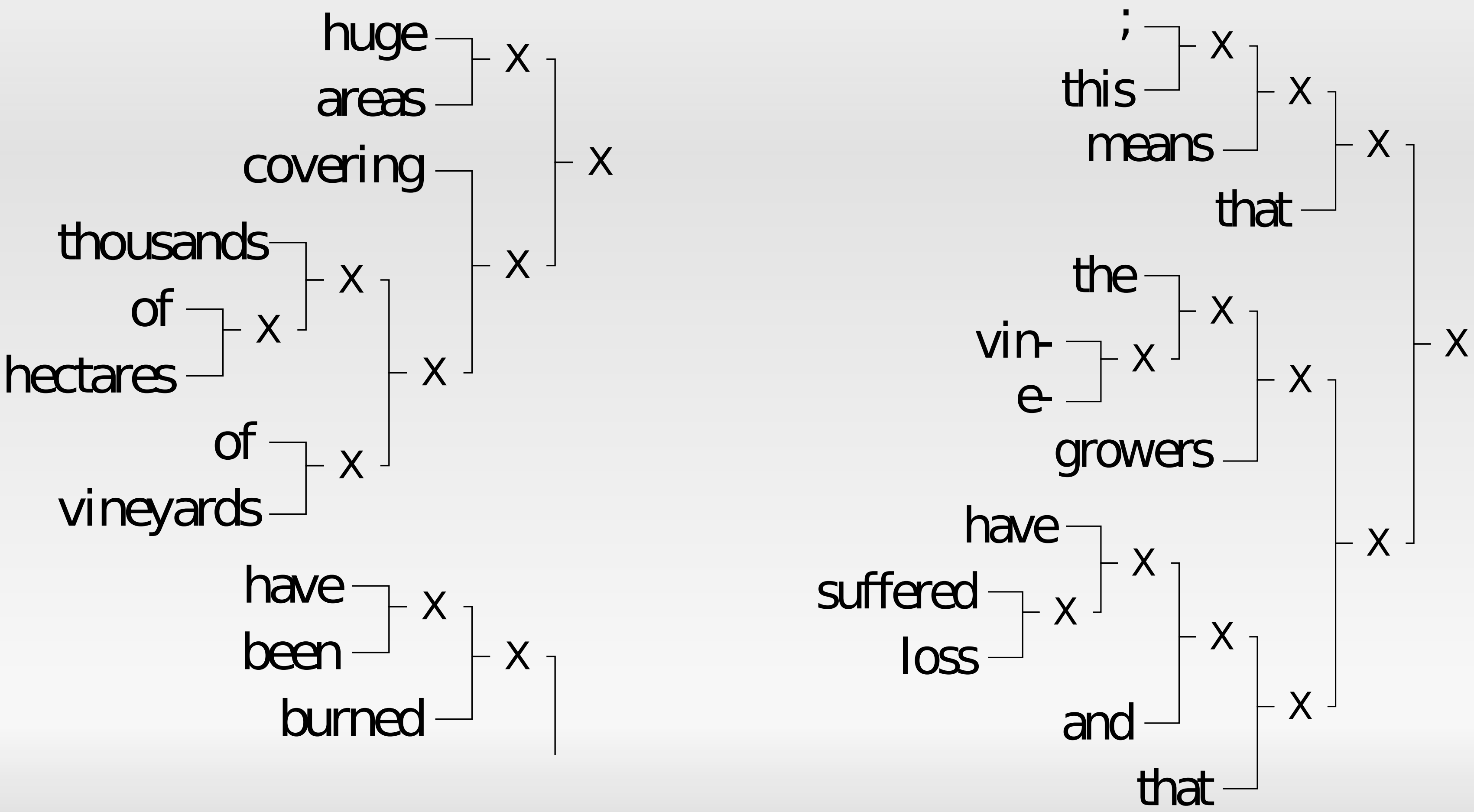


s(T) = s(ab) + s(abc)

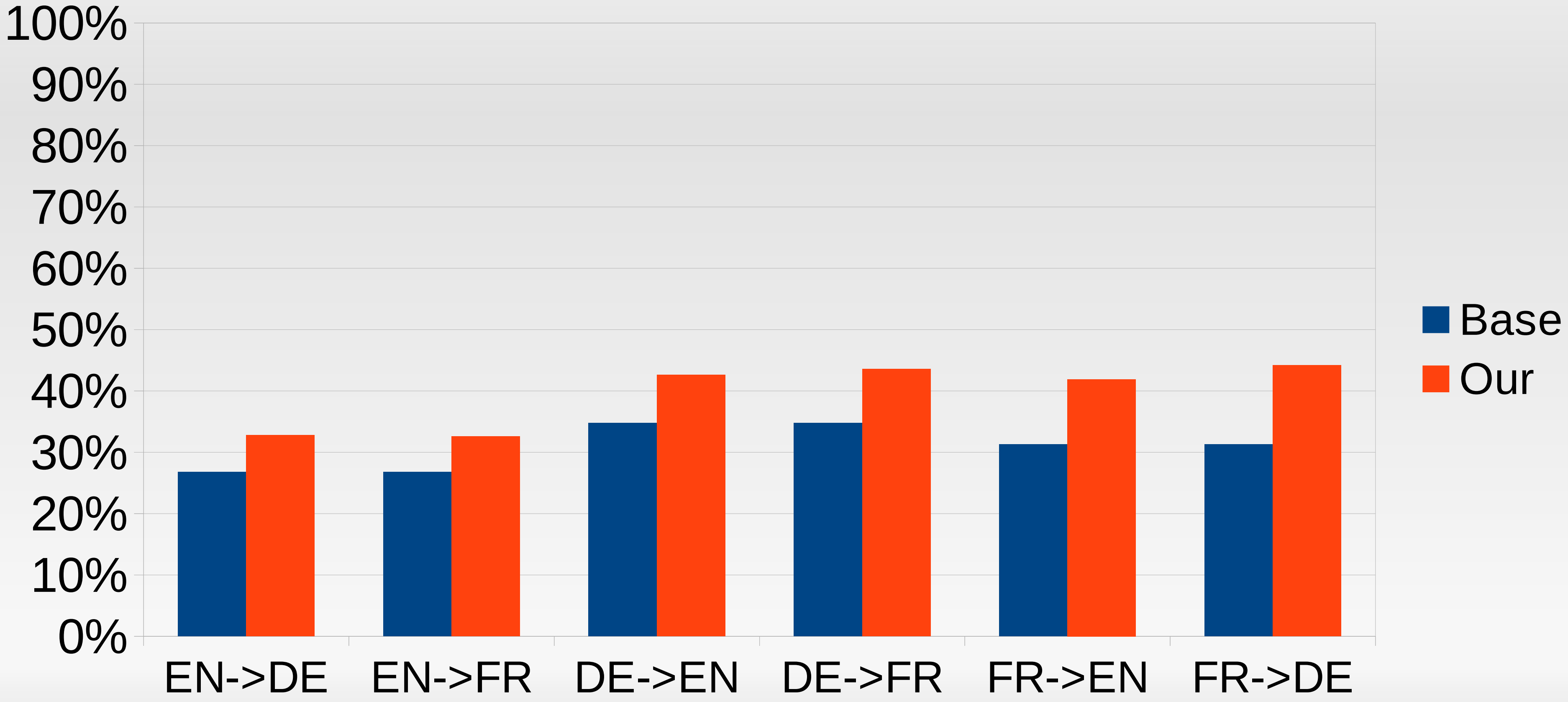
Results


X
huge
areas
covering
thousands
of
hectares
of
vineyards
have
been
burned
;
this
means
that
the
vin-
e-
growers
have
suffered
loss
and
that
their
plants
have
been
damaged
.
EOS

Results


huge
areas
covering
thousands
of
hectares
of
vineyards
have
been
burned
X
;
this
means
that
the
vin-
e-
growers
have
suffered
loss
and
that
X

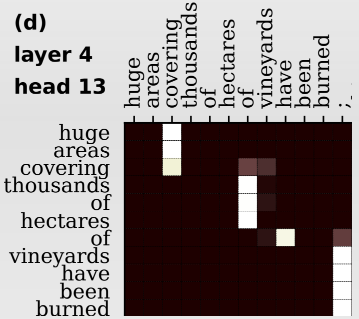
Comparison to standard syntactic trees



Summary

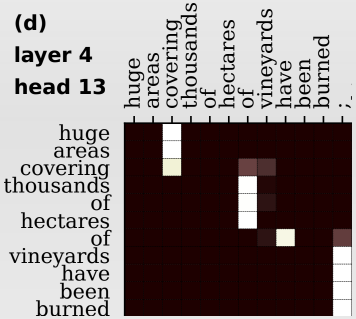
Summary

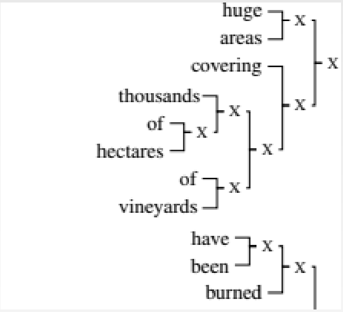
- **Balusters** in Transformer NMT encoder self-attentions
  - Contiguous sequence of output states
  - Attention to the same one input state



Summary

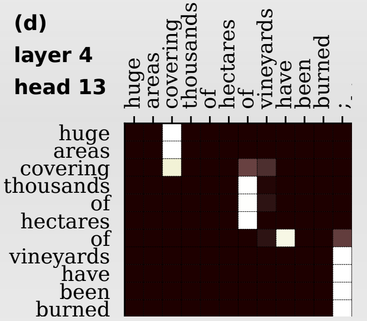
- **Balusters** in Transformer NMT encoder self-attentions
  - Contiguous sequence of output states
  - Attention to the same one input state
- Interpret balusters as **syntactic phrases**
  - Phrase candidate extraction and scoring
- Construct a binary **constituency tree**
  - CKY algorithm

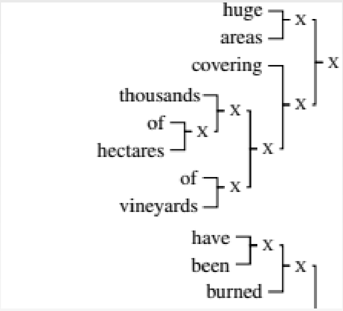




Summary

- **Balusters** in Transformer NMT encoder self-attentions
  - Contiguous sequence of output states
  - Attention to the same one input state
- Interpret balusters as **syntactic phrases**
  - Phrase candidate extraction and scoring
- Construct a binary **constituency tree**
  - CKY algorithm
- Compare to **standard syntactic trees**
  - **~40% match**; base ~30% match





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/lsd