Experiments on human incremental parsing

Leonid Mityushin Leonid Iomdin

Laboratory of Computational Linguistics

Institute for Information Transmission Problems

Russian Academy of Sciences

ETAP – multilingual multifunctional linguistic processor

dependency parsing, machine translation, semantic analysis, question answering

Languages

Russian (~110,000 word lexicon)
English (~90,000 word lexicon)
French, German, Spanish, Korean, Arabic

SynTagRus – Russian dependency treebank ~1,100,000 words

Incremental text comprehension

At any moment, the reader/listener has a complete or almost complete linguistic and pragmatic interpretation of the part of the text perceived up to that moment. This interpretation, as a rule, does not change after new parts of the text have been perceived.

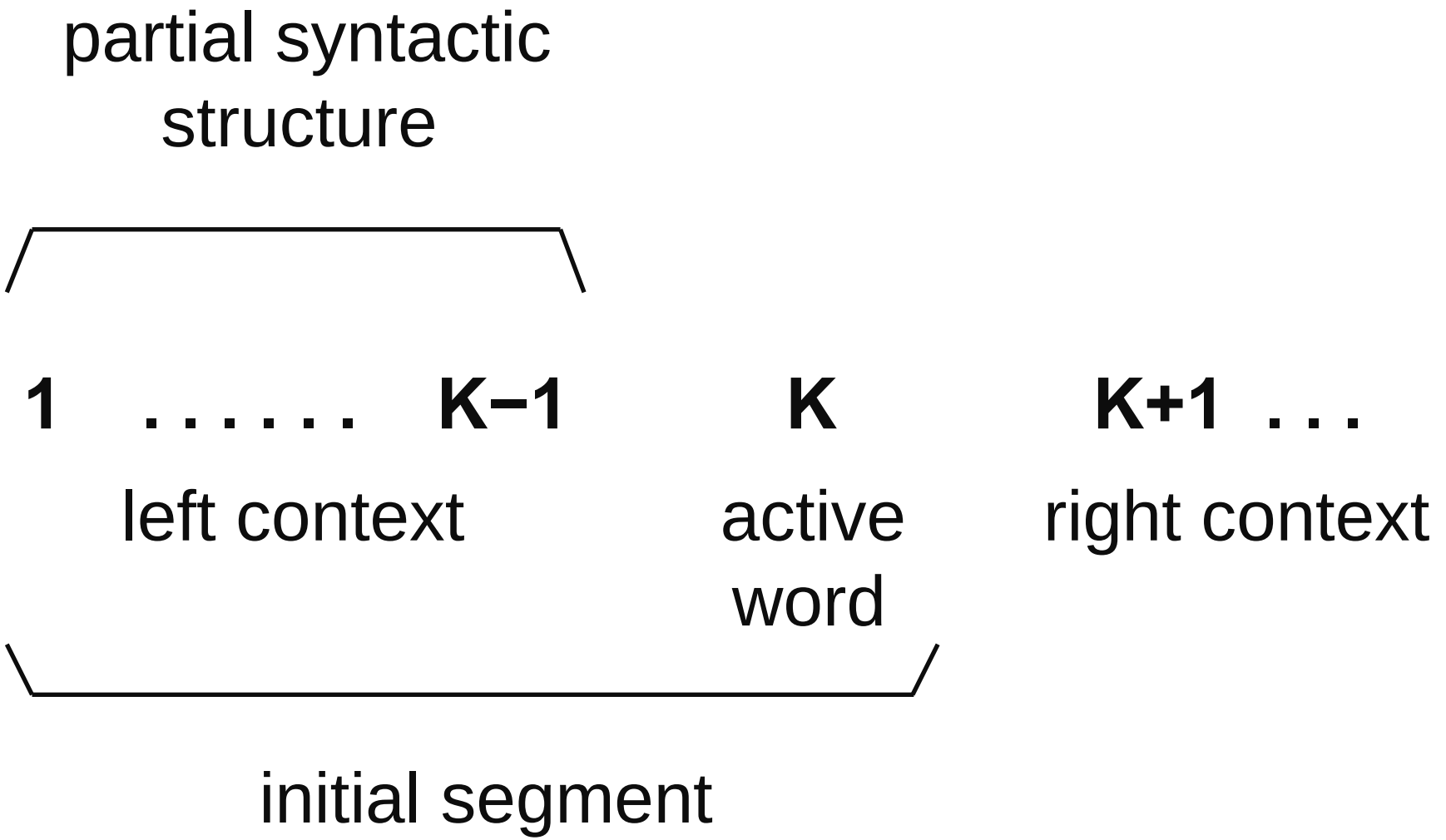
The aim of this work is to evaluate whether this is true for human comprehension of the syntactic structure of a sentence.

ETAP syntactic model

Nodes of a dependency tree = words of the sentence (not punctuation marks).

Syntactic links = directed arcs between words, labelled with names of syntactic relations.

Russian syntax: about 70 syntactic relations.


partial syntactic structure
1 . . . . . K−1
K
K+1 . . .
left context
active word
right context
initial segment

100 percent confident incremental parsing is impossible.

Size of the right context = 0
Initial segment: ***John met her . . . .*** – ???

John met her yesterday.

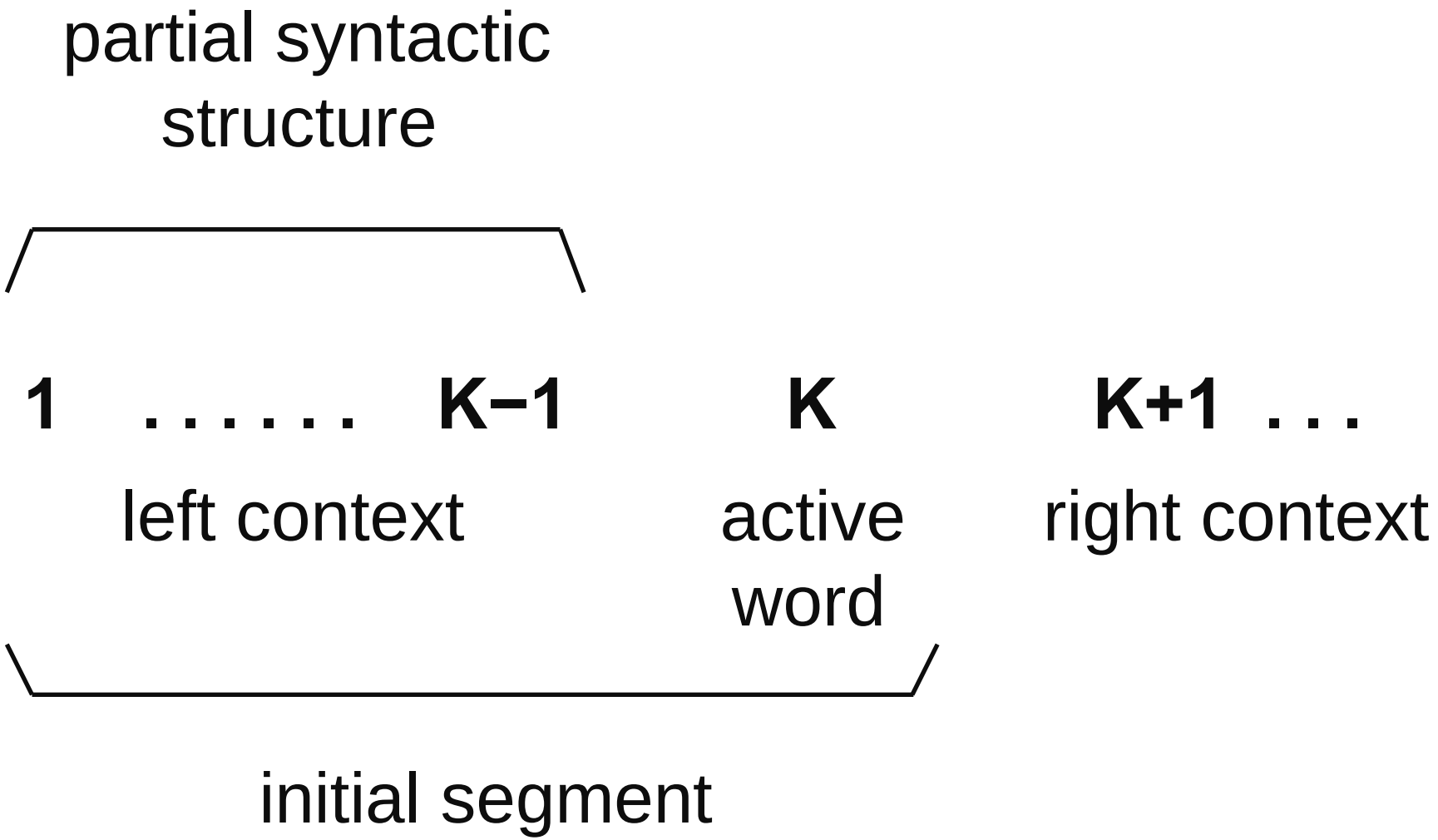
John met ⟶ her yesterday

John met her sister yesterday.

John met her ⟵ sister yesterday

Tentative links

Given the initial segment ***John met her . . . .*** (with a zero right context), we create a syntactic link *met* ⟶ *her* but mark it as **tentative** (the other links are called **final**).


partial syntactic structure
1 . . . . . K−1
K
K+1 . . .
left context
active word
right context
initial segment

We presume that processing a sentence always results in creating its correct complete dependency tree.

Performance indicators

number of corrections

number of created tentative links

London Orbital is a 117 mile long motorway, encircling almost all of Greater London.

size of the right context = 1

London **Orbital** is .....

1 London --
2 **Orbital**
is

```
-------------------------------
 * -->  2              |
 2 -->  1  London      |
-------------------------------
```

```
TENTATIVE LINKS
-------------------------------------
create and insert |    -->
create            |    -->
insert            |    -->
remove            |    -->
-------------------------------------

CORRECTION OF FINAL LINKS
-------------------------------------
insert            |    -->
remove            |    -->
-------------------------------------
```

London **Orbital** is .....

1 London --
2 **Orbital**
is

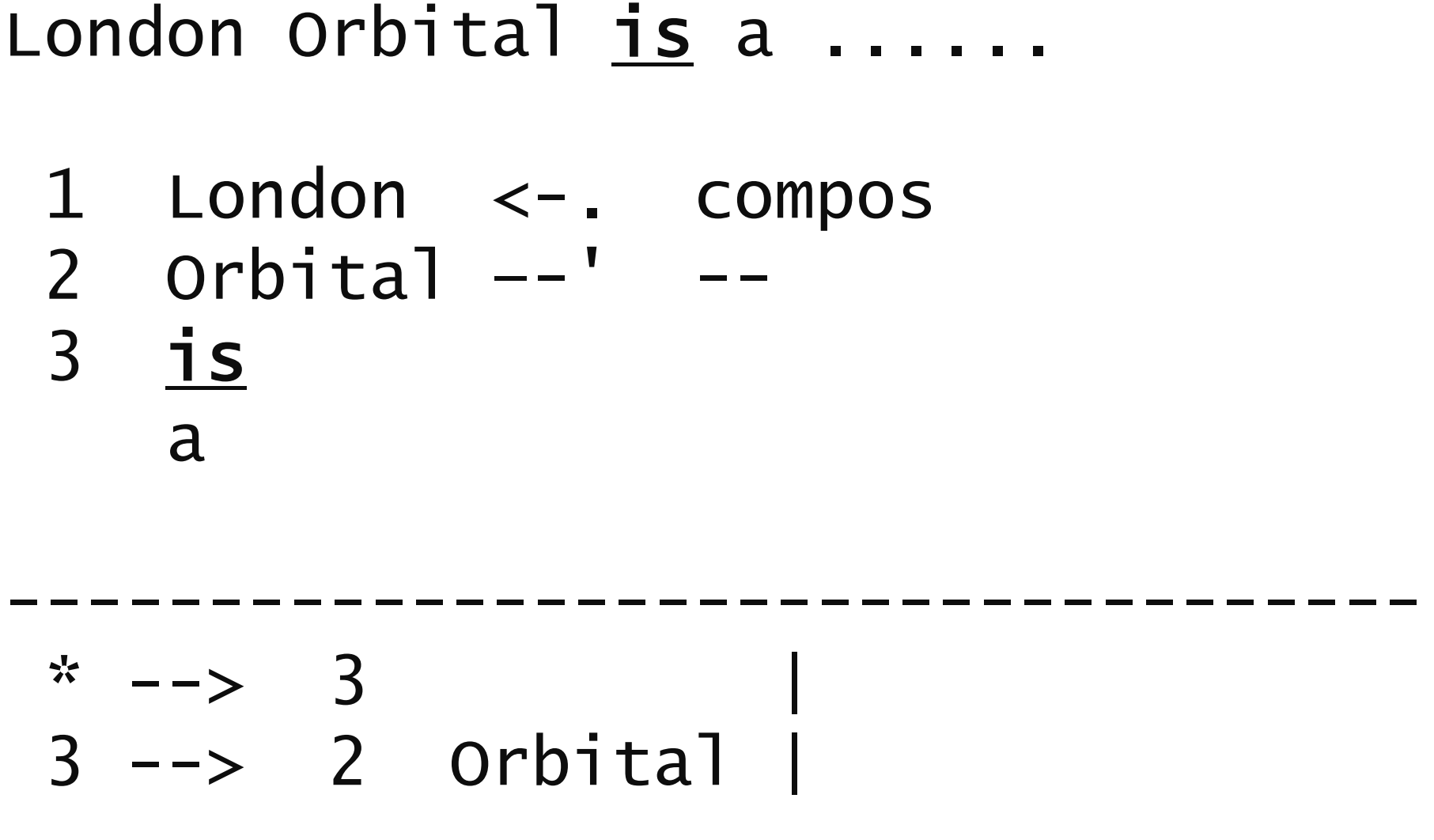
```
-------------------------------
 * -->  2              |
 2 -->  1  London      |
-------------------------------
```

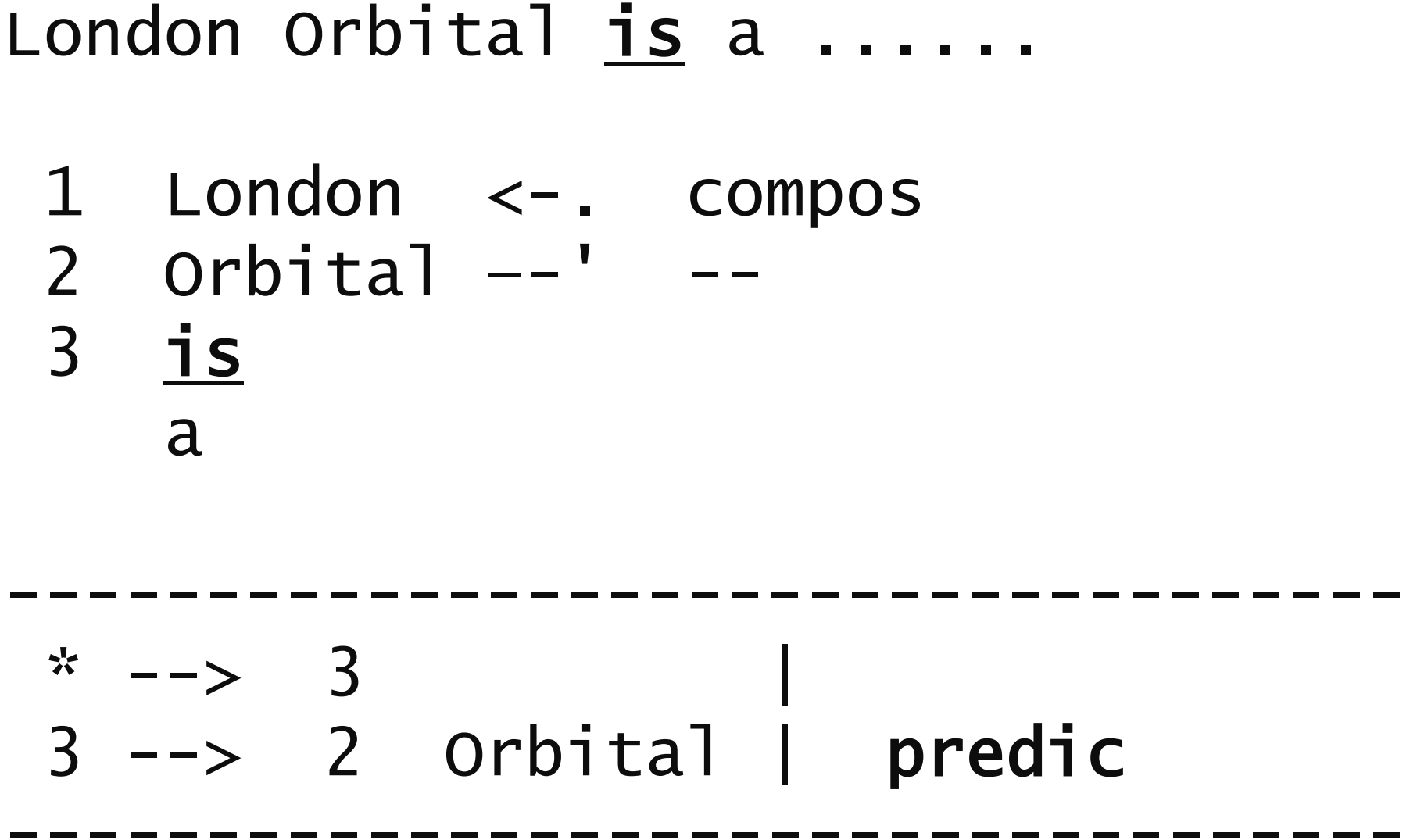
London **Orbital** is .....

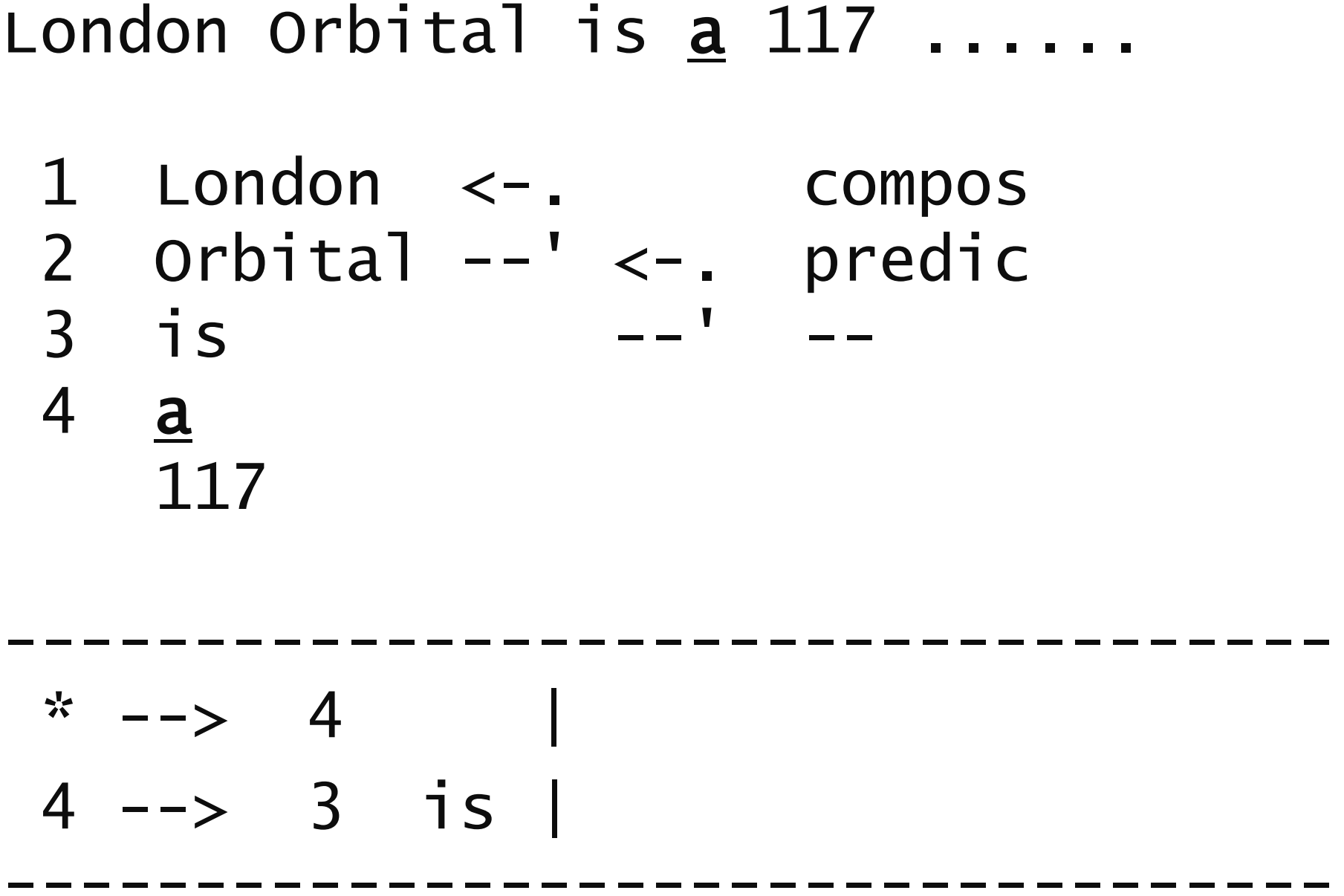
1 London --
2 **Orbital**
is

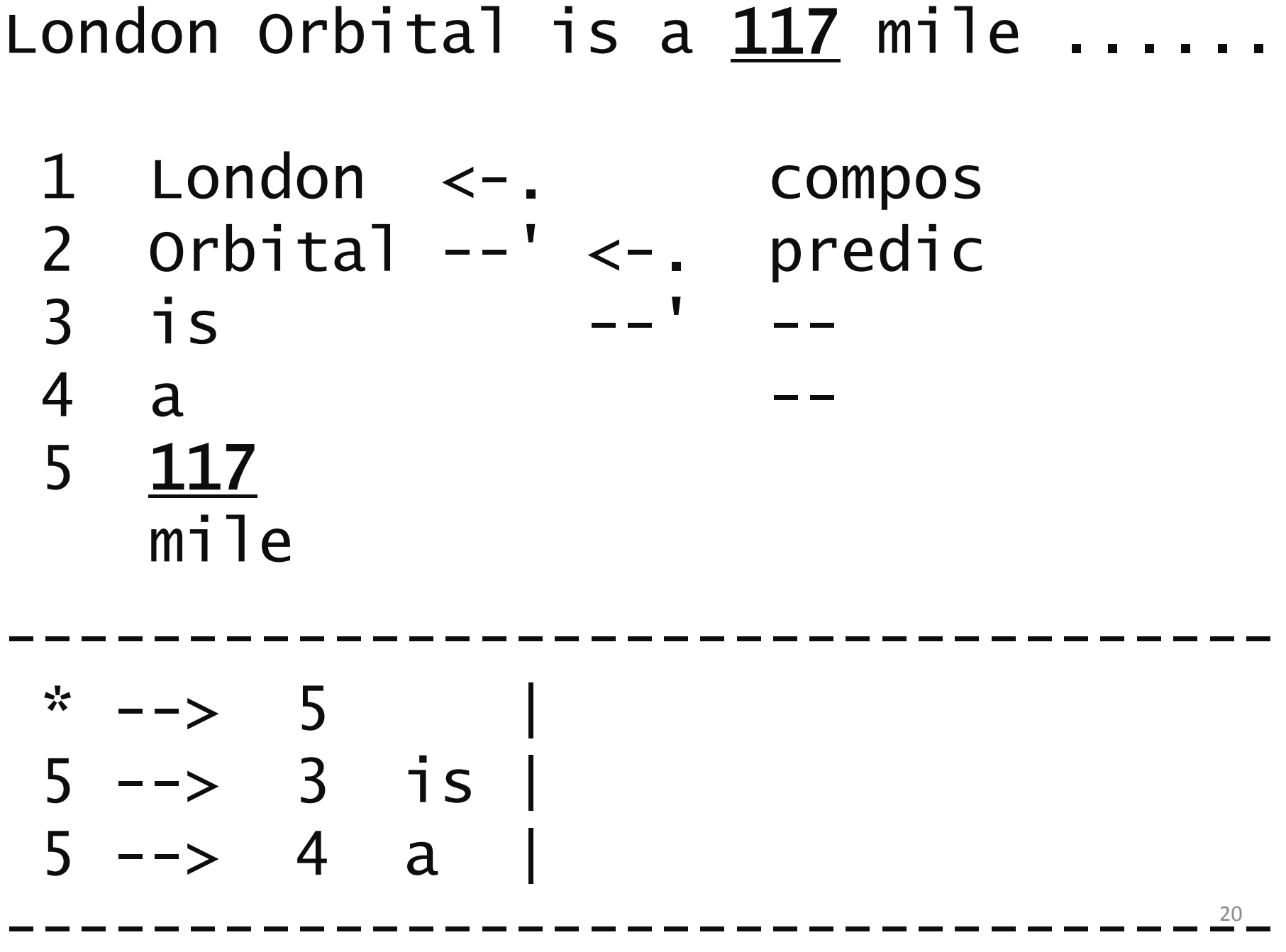
---

* --> 2 |
2 --> 1 London | **compos**


London Orbital is a .....
1 London <-. compos
2 Orbital --' --
3 is
a
* --> 3
3 --> 2 Orbital


London Orbital is a .....
1 London <-. compos
2 Orbital --' --
3 is
a
* --> 3 |
3 --> 2 Orbital | predic


London Orbital is a 117 .....
1 London <-. compos
2 Orbital --' <-. predic
3 is --' --
4 a
117
* --> 4 |
4 --> 3 is |


London Orbital is a 117 mile .....
1 London <-. compos
2 Orbital --' <-. predic
3 is --' --
4 a --
5 117
mile
* --> 5 |
5 --> 3 is |
5 --> 4 a |

```
London Orbital is a 117 mile long
......

 1  London  <-.       compos
 2  Orbital --' <-.   predic
 3  is          --'   --
 4  a                 --
 5  117               --
 6  mile
    long
```

London Orbital is a 117 **mile** long

```
......

-----------------------------------
 *  -->  6       |
 6  -->  3  is   |
 6  -->  4  a    |
 6  -->  5  117  |
-----------------------------------
```

London Orbital is a 117 **mile** long

```
......

-----------------------------------
 *  -->  6          |
 6  -->  3   is     |
 6  -->  4   a      |
 6  -->  5   117    |   quantit
-----------------------------------
```

London Orbital is a 117 mile **long** motorway .....

```
 1  London  <-.       compos
 2  Orbital --' <-.   predic
 3  is          --'   --
 4  a                 --
 5  117     <-.       quantit
 6  mile    --'       --
 7  long
    motorway
```

London Orbital is a 117 mile **long** motorway .....

```
-----------------------------------
 *  -->  7          |
 7  -->  3  is      |
 7  -->  4  a       |
 7  -->  6  mile    |
-----------------------------------
```

London Orbital is a 117 mile **long** motorway .....

```
-----------------------------------
 *  -->  7          |
 7  -->  3  is      |
 7  -->  4  a       |
 7  -->  6  mile    |  restr
-----------------------------------
```

London Orbital is a 117 mile long **motorway**, encircling .....

```
1  London      <-.        compos
2  Orbital     --' <-.    predic
3  is              --'    --
4  a                      --
5  117         <-.        quantit
6  mile        --' <-.    restr
7  long            --'    --
8  motorway,
   encircling
```

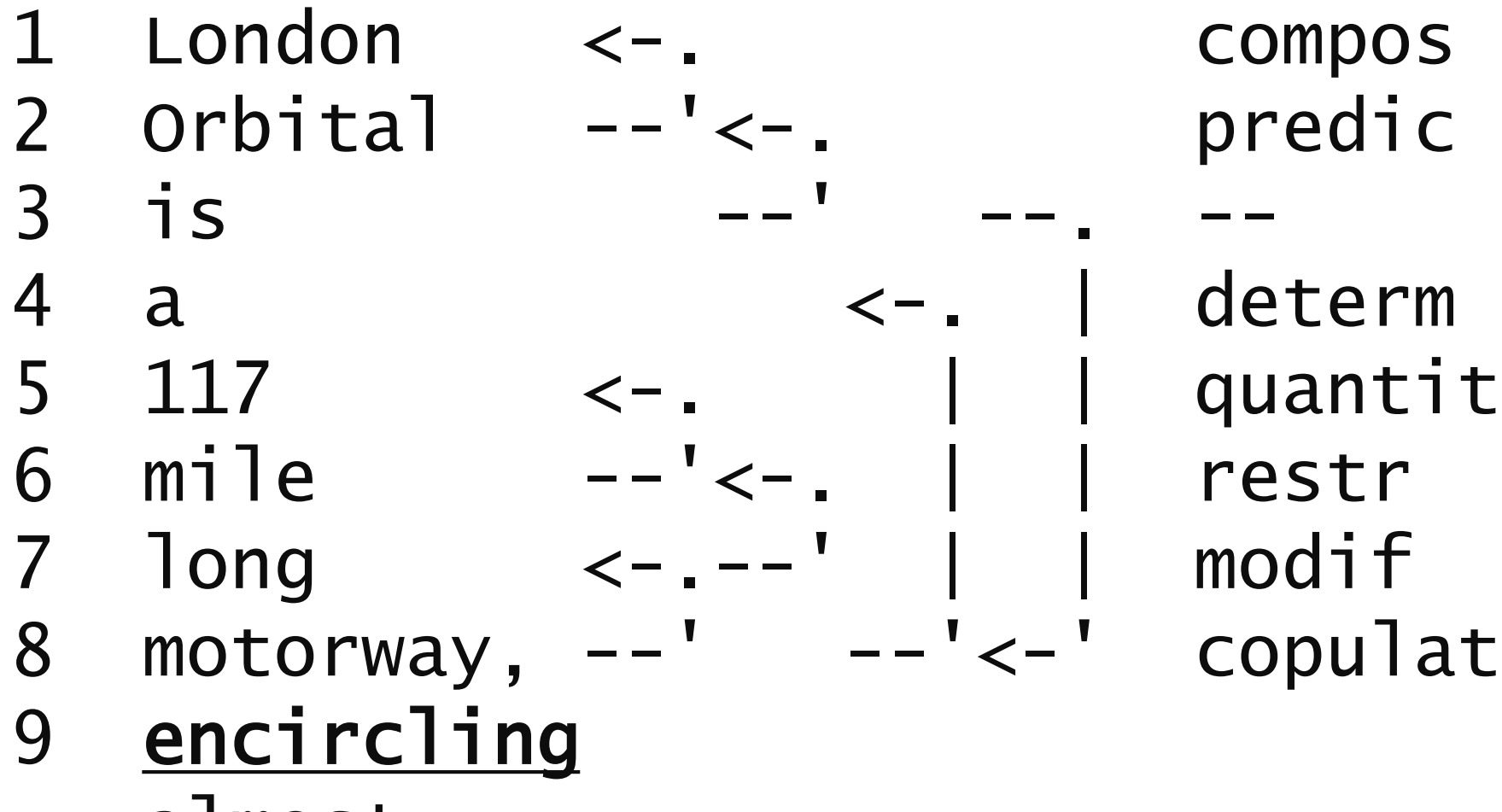
London Orbital is a 117 mile long **motorway**, encircling .....

```
-----------------------------------
 * -->  8        |
 8 -->  3  is    |
 8 -->  4  a     |
 8 -->  7  long  |
-----------------------------------
```

London Orbital is a 117 mile long **motorway**, encircling .....

*	-->	8		3 **copulat**
8	-->	3	is	
8	-->	4	a	**determ**
8	-->	7	long	**modif**

London Orbital is a 117 mile long motorway, **encircling** almost .....



London Orbital is a 117 mile long motorway, **encircling** almost .....

```
-------------------------------
 * -->  9                |
 9 -->  8  motorway |
-------------------------------
```

London Orbital is a 117 mile long motorway, **encircling** almost .....

```
-------------------------------------
 * -->  9                |  8  modif
 9 -->  8  motorway |
-------------------------------------
```

and so on ...

Three series of experiments were conducted for the sizes of the right context 0, 1 and 2, with 100 sentences processed in each series.

The role of the subjects was played by the authors of this paper.

The sentences for the experiments were selected at random from the two sets of sentences offered as training material for the competition "Automatic Gapping Resolution for Russian“. Only non-elliptical sentences were used.

right context	total number of links	tentative links in the trees	created tentative links	number of correc-tions
0	1627	**34** = 2.2%	**75** = 4.6%	**3**
1	1741	**21** = 1.2%	**34** = 2.0%	**0**
2	1607	**8** = 0.5%	**13** = 0.8%	**0**

Thank you!