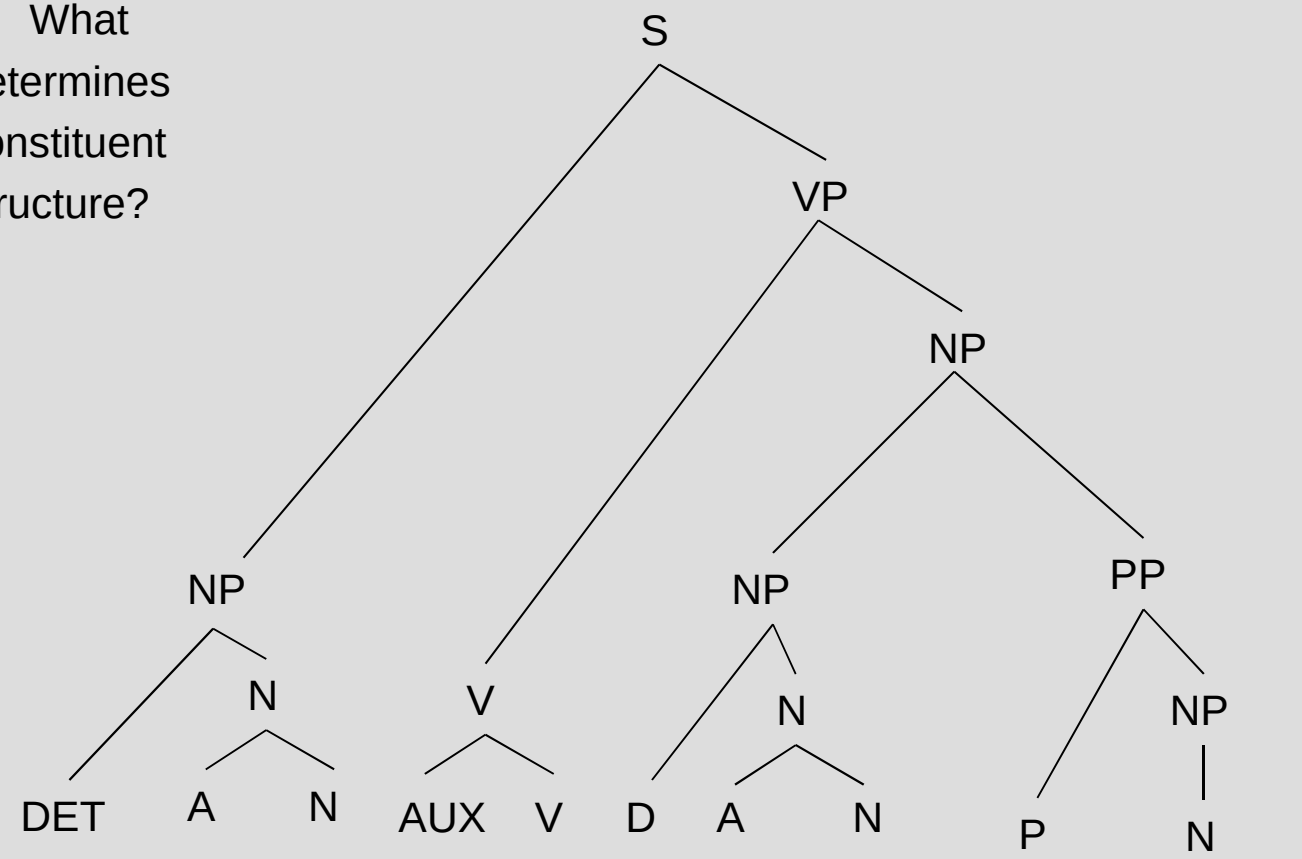
Prefabs and constituent structure

Holger Diessel
holger.diessel@uni-jena.de

Bybee and Scheibman 1999

- What determines constituent structure?



Bybee and Scheibman 1999

- What determines constituent structure?

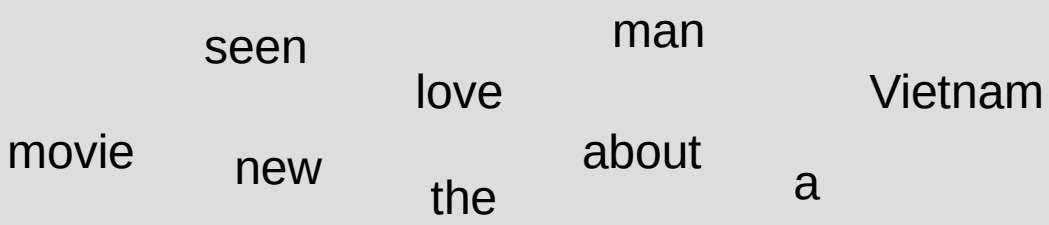
  Two factors:

  - Meaning
  - Frequency

  The importance of meaning for linear order and constituency has been well-known since Behaghel (1932)

> *Geistig eng Zusammengehöriges wird auch eng zusammengestellt*.
> 'Conceptually related entities are placed close to each other.'
> [Behaghel's First Law]

Bybee and Scheibman 1999



The young man has seen a new movie about Vietnam

Bybee and Scheibman 1999

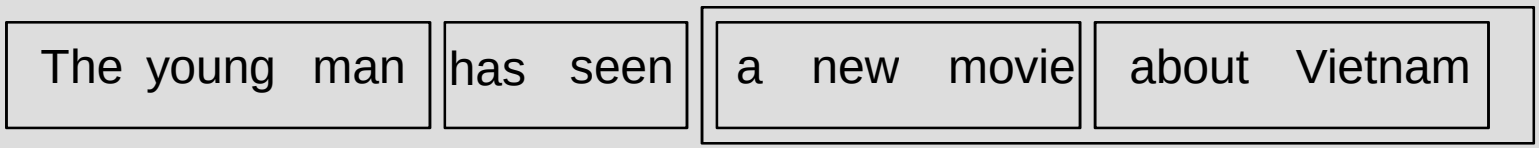
seen man

love Vietnam

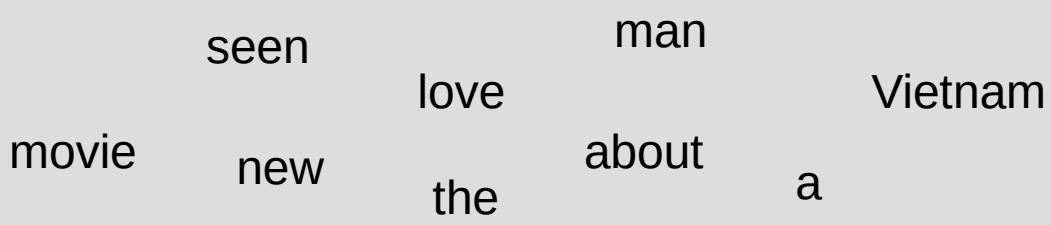
movie about

new a

the



Bybee and Scheibman 1999



Man young the Vietnam about movie new a seen has

Bybee and Scheibman 1999

seen man

love Vietnam

movie about

new a

the

Bybee and Scheibman 1999

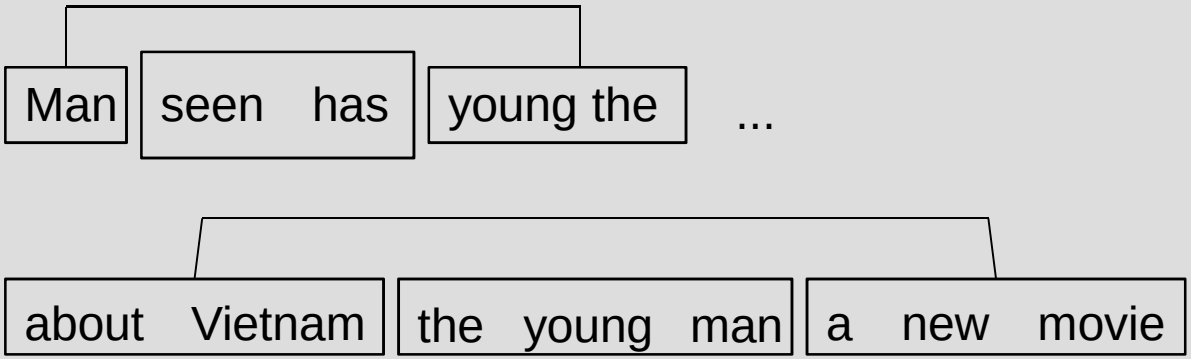
seen man

love Vietnam

movie about

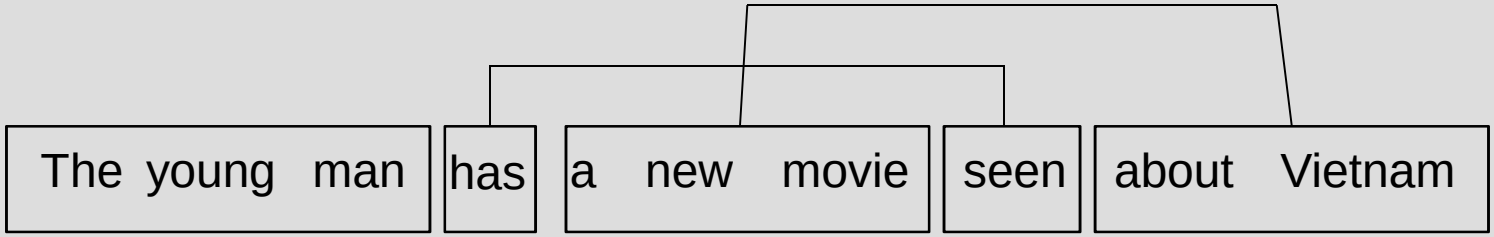
new a

the



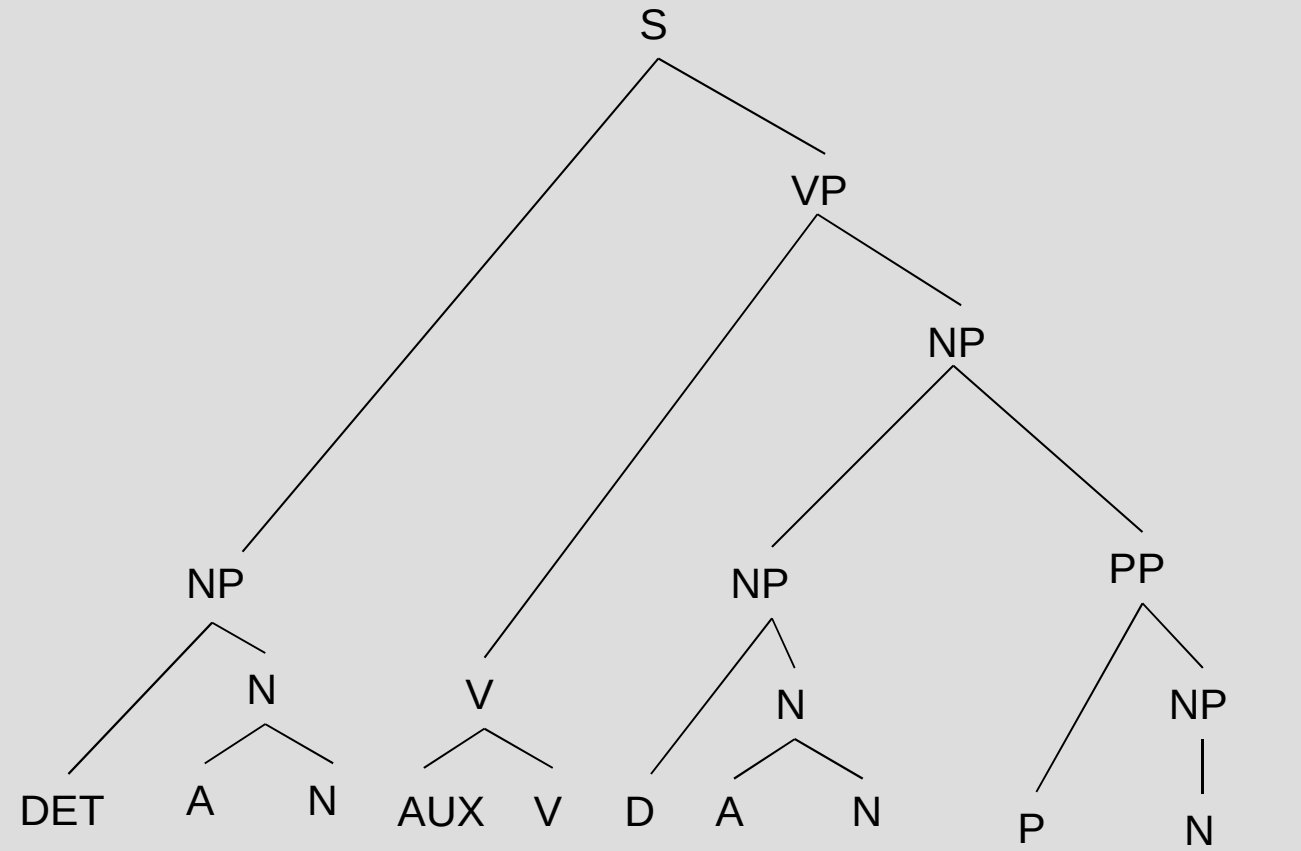
Bybee and Scheibman 1999

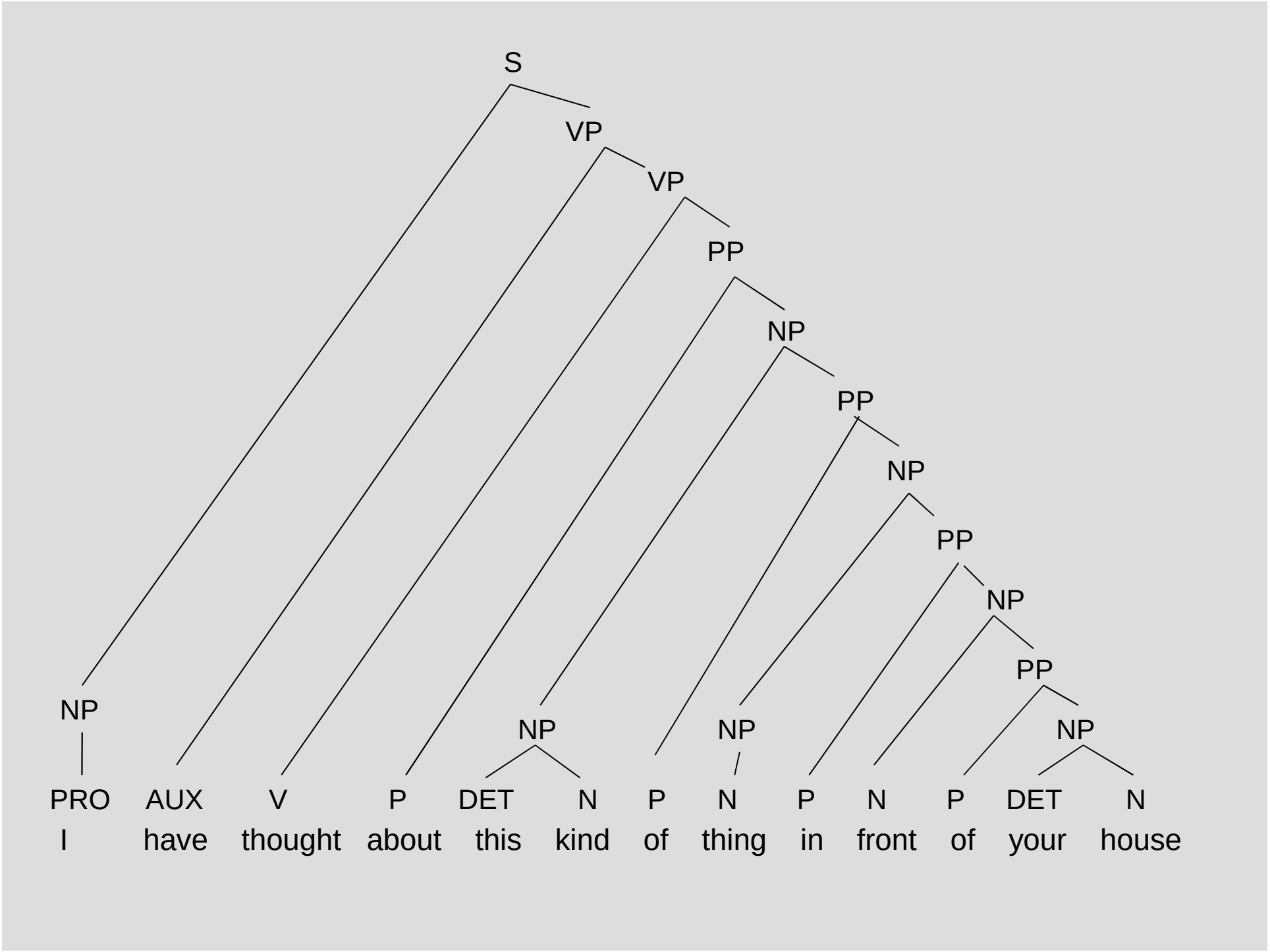
But there are ‘strange’ languages:

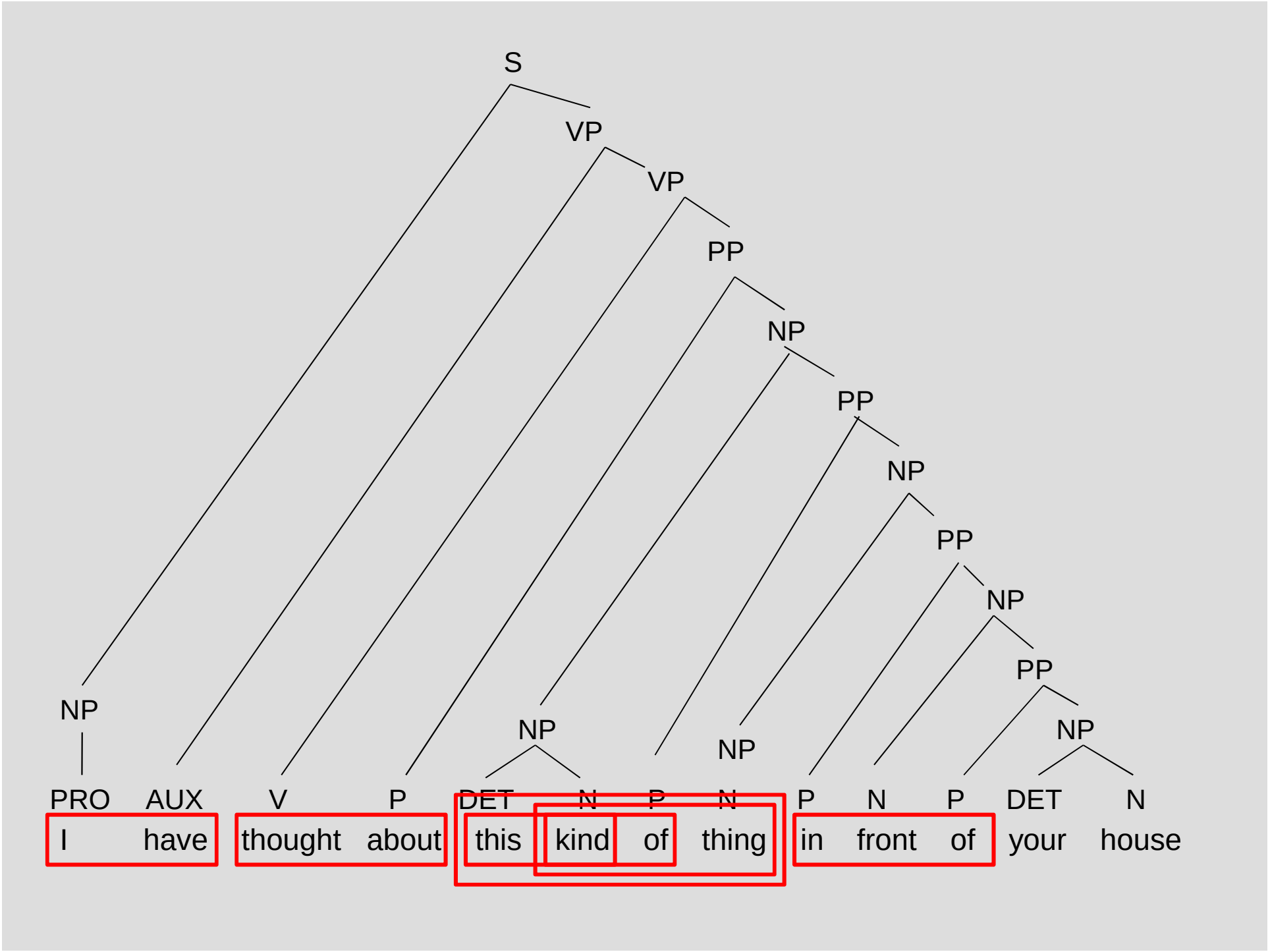


The horrors of the German language

Constituent structure tree




S
VP
VP
PP
NP
PP
NP
PP
NP
PP
NP
NP
NP
NP
PRO
AUX
V
P
DET
N
P
N
P
N
P
DET
N
I
have
thought
about
this
kind
of
thing
in
front
of
your
house

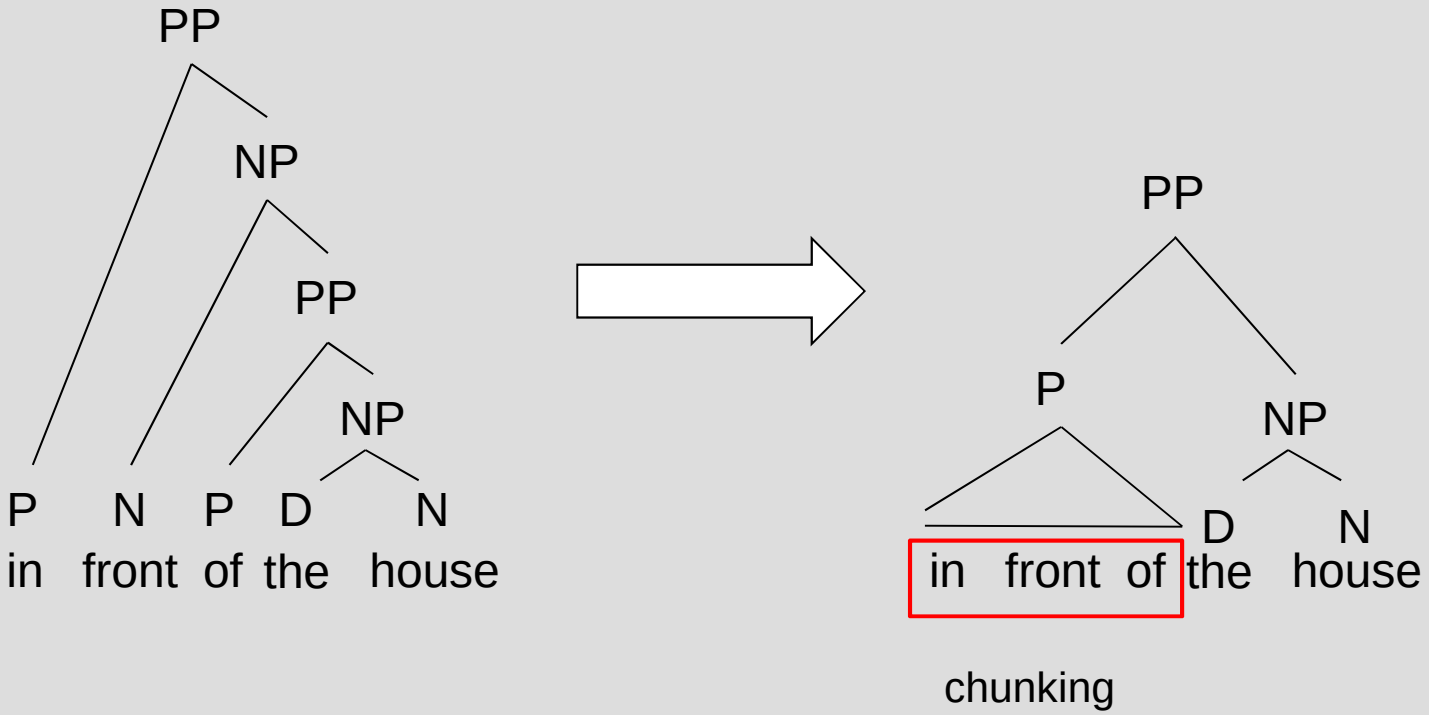

S
VP
VP
PP
NP
PP
NP
PP
NP
PP
NP
NP
NP
NP
PRO
AUX
V
P
DET
N
P
N
P
N
P
DET
N
I
have
thought
about
this
kind
of
thing
in
front
of
your
house

Grammaticalization

Grammaticalization can change constituent structure:

Grammaticalization

Grammaticalization can change constituent structure:

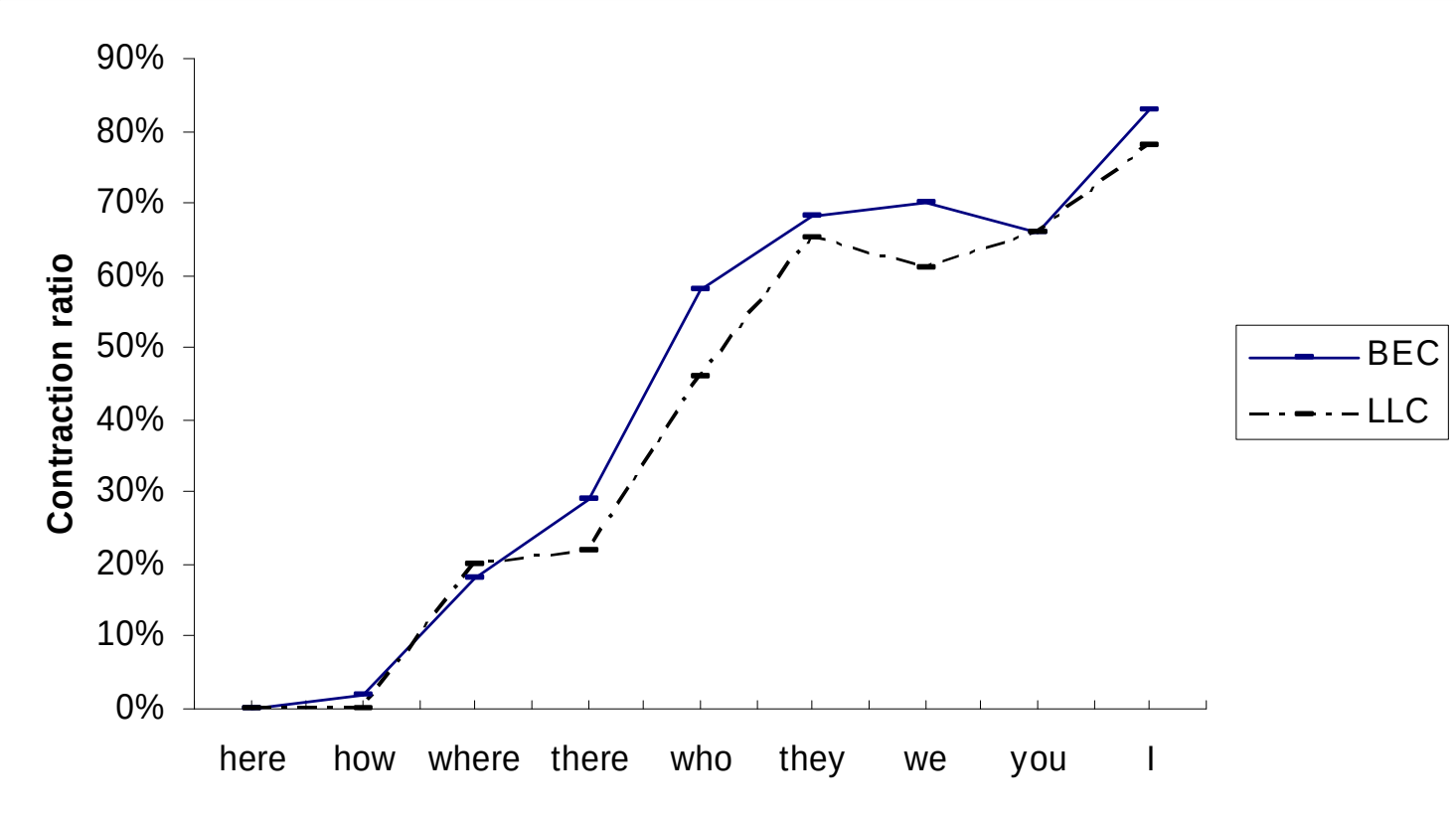


Krug 1998

The reduction of auxiliaries.

that is	vs.	that’s
we will	vs.	we’ll
I have	vs.	I’ve

Krug 1998


90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Contraction ratio
here
how
where
there
who
they
we
you
I
BEC
LLC

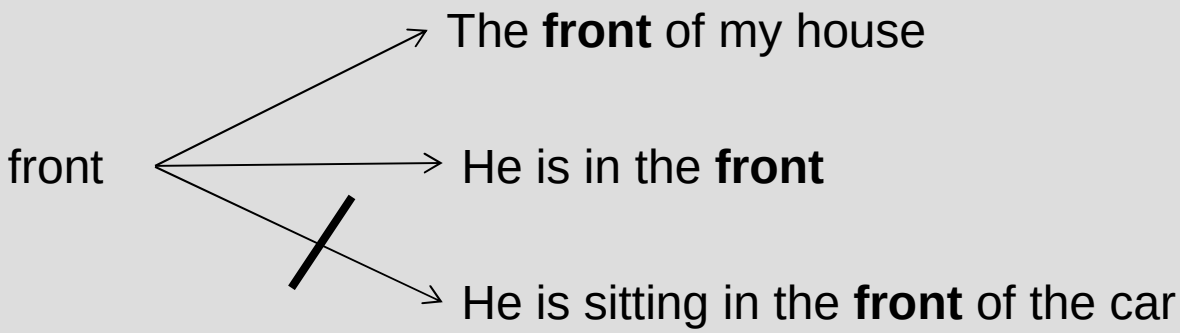
The linear fusion hypothesis

Bybee (2002): “The the Linear Fusion Hypothesis”

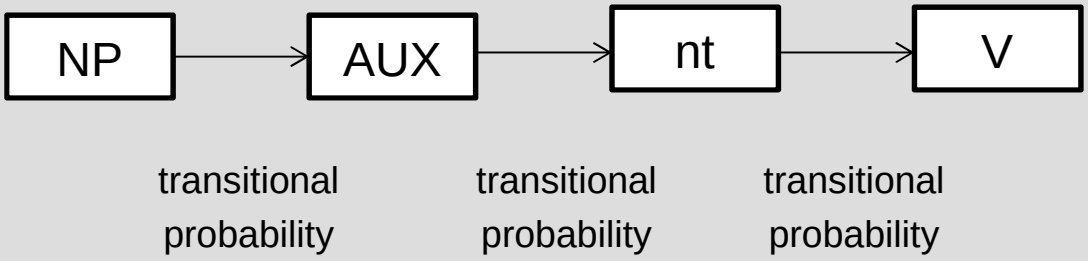
Items that are used together fuse together.

Autonomy

Bybee (2002): “Chunks become autonomous”



The reduction of *don't*



$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Transitional probability

In a corpus including 12.000 nouns and 3.500 adjectives, 2.000 adjectives precede a noun.

(1) What is the likelihood that an noun follows an adjective?

$$P(ADJ \mid N) = \frac{P(ADJ \cap N)}{P(N)}$$

$$P(ADJ \mid N) = \frac{P(2000)}{P(12000)} = 0.1666$$

The reduction of *don't*

- Tokens with an initial [d] and a full vowel [dõt, dõn]
- Tokens with an initial flap and a full vowel [rõt, rõ]
- Tokens with a flap and a reduced vowel [rˇə]
- Tokens with just a reduced vowel [rˇə, ə]

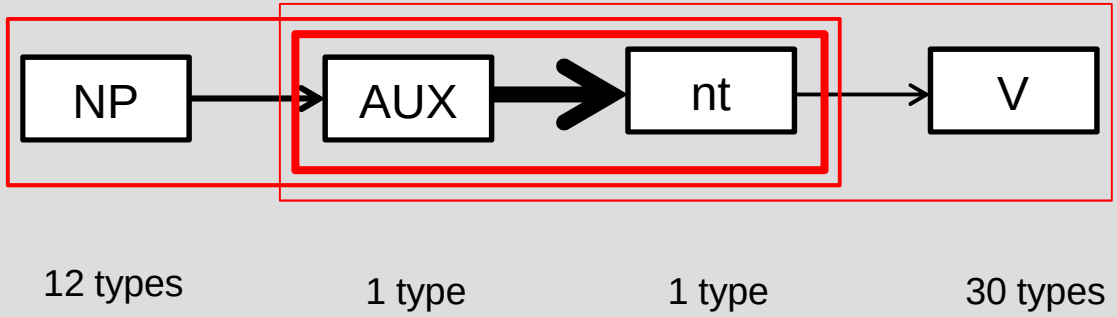
The reduction of *don't*

Preceding NP	*dõt, dõn, dõ*	*rõt, rõ*	*rə̃*	*ə̃, ə*	Total	Percentage
I	16	22	38	12	88	63%
you	7	7	-	-	14	10%
we	2	6	-	-	8	6%
they	1	3	-	-	4	3%
Lexical NP	5	-	-	-	5	4%
...	...	...	...	...	...	...
Total	43	44	39	12	138	100%

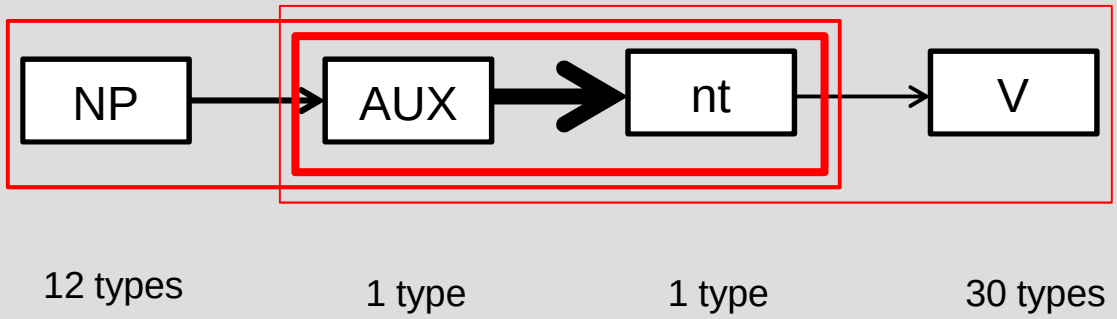
The reduction of *don't*

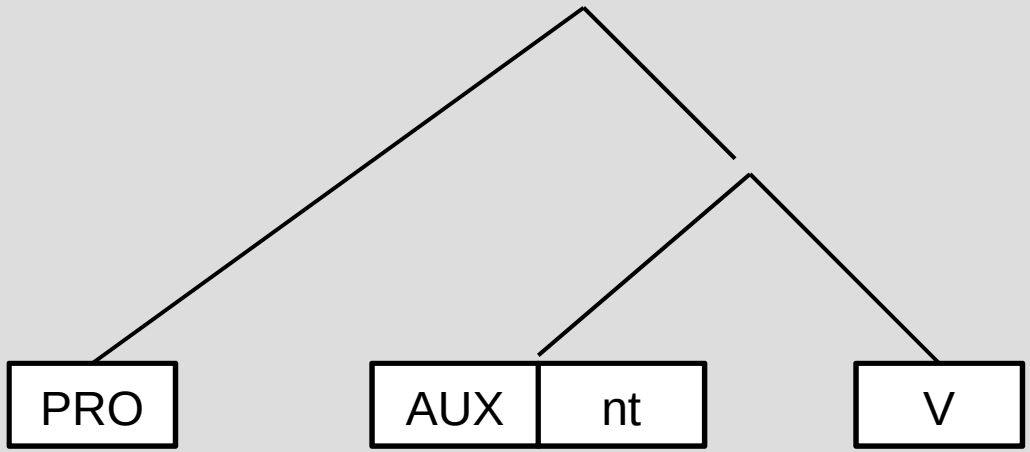
Following V	*dõt, dõn, dõ*	*rõt, rõ*	*rə̃*	*ə̃, ə*	Total	Percentage
know	2	8	24	5	39	36.8%
think	7	6	6	1	20	18.9%
have	1	7	1	-	9	8.5%
have to	1	2	1	-	4	3.8%
want	1	1	3	-	5	4.7%
see	3	1	-	-	4	3.8%
like	-	1	-	1	2	1.9%
get	1	2	-	-	3	2.8%
mean	-	-	-	1	1	0.9%
feel	-	-	-	1	1	0.9%
...	...	...	...	...	...	...
Verb tokens	25	36	36	9	106	100%

The reduction of *don't*

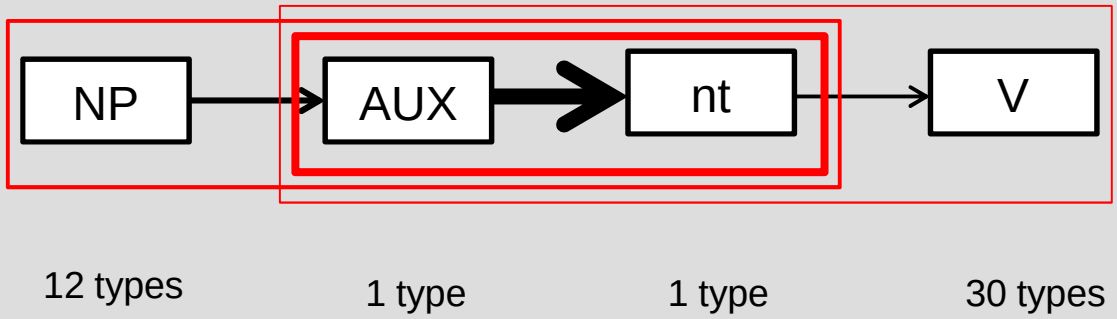


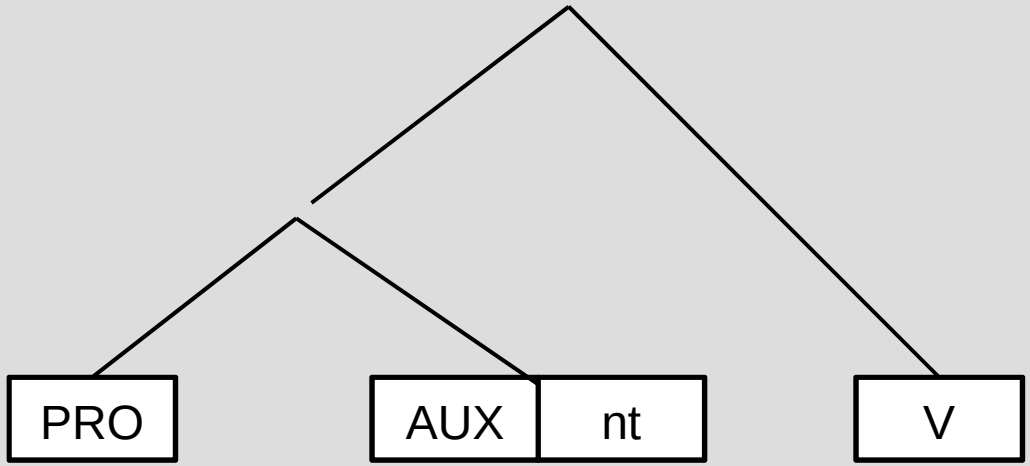
The reduction of *don't*





The reduction of *don't*





Conflicts between meaning and frequency

Since there is a higher transitional probability between subject pronouns and *don‘t* than between *don‘t* and the subsequent verb, the grouping by meaning and the grouping by frequency are in conflict with each other.

Usually meaning and frequency are in unison:

NP → Det N		
the	+	N
a	+	N
some	+	N

VP → AUX V		
have	+	V-ed
will	+	V-inf
must	+	V-inf

But sometimes they are in conflict with each other:

Pro → don’t → V				
I	+	don’t	+	V
I	+	don’t	+	V
Why	+	don’t	+	you

Conflicts between meaning and frequency

Other examples of conflict:

P-DET-N				
an	+	dem	+	N
zu	+	dem	+	N
auf	+	dem	+	N

V-P-N				
think	+	of	+	N
talk	+	about	+	N
dream	+	of	+	N

I don‘t know

(1) … because I was telling F, ***I don’t know*** any woman that I’ve discussed it with, who hasn’t tried it.

(2) Z: well I talked to a guy that’s thirty-four in my class, and we were talking about the difference, just in … phonology and how you feel, and your best study hours, and [the …] rest you need versus what you do,
O: [right]
Z: ***I don’t [rə] know***
It’s just –
O: right
S: you know what was the biggest give away for me?
I couldn’t –
I can’t get loaded anymore.
Z: uh huh
All: @@@@
S: I mean without paying.

I don't think

(1) It's really horrible.
so … if I take this stuff
I'll drink it but,
I don't think about taking it.

(2) ZS: you guys need some capital, I keep saying.
F: we need the Mormon church behind us.
O: Oh?
S: A well you get them.
F: How?
O: Yeah **I don't think** they'll go for your fantasy.

Why don‘t you

(1) I really don’t know what you mean.

Why don’t you want to help your brother?

(2) S: She asked me a question
I say
no that’s not one question,
so I started telling her how it’s …
Z: hmm
S: more than one question.
I said **why don’t you** sit down,
so that I can talk to you about it.

Conclusion

Constituent structures of the type proposed for generative grammar that are described by phrase structure trees do not exist. Instead, units of language are combined into chunks as a result of frequent repetition.

Most of the time these chunks bear a semantic and/or pragmatic relation to one another allowing them to fulfill the grammatical criteria for constituency: they can be used alone; they can be replaced by a pro-form; etc.

But some chunks are in conflict with semantic groups and traditional constituents: pro-aux, V-P.

On this view constituent structure is emergent and gradient: it varies with the transitional frequencies between individual lexical expressions.