Algebraic decoder specification: coupling formal-language theory and statistical machine translation

Thesis defense
translated from German into English

Matthias Büchse

TECHNISCHE UNIVERSITÄT DRESDEN

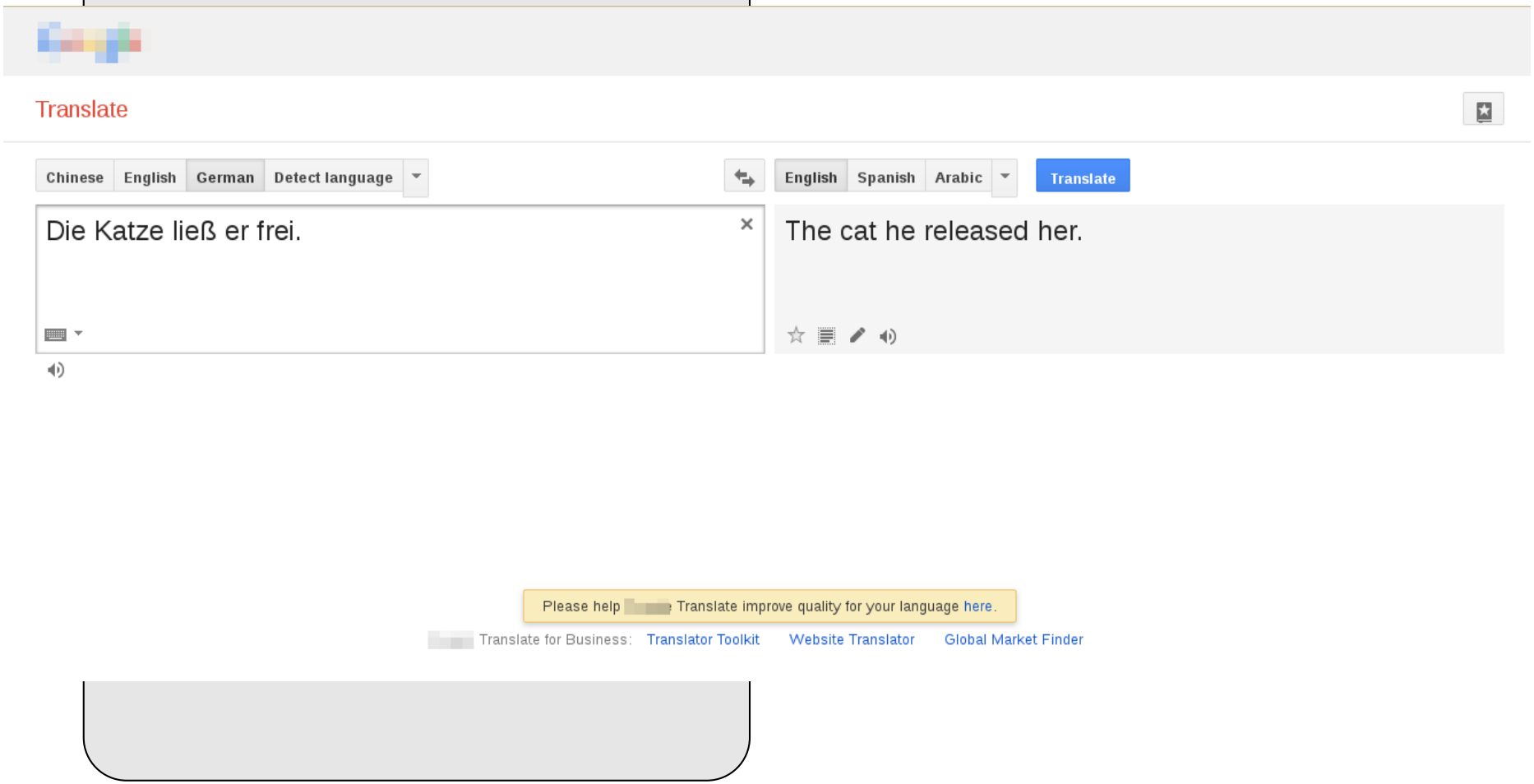
2014-12-18

Overview

Statistical
Machine Translation

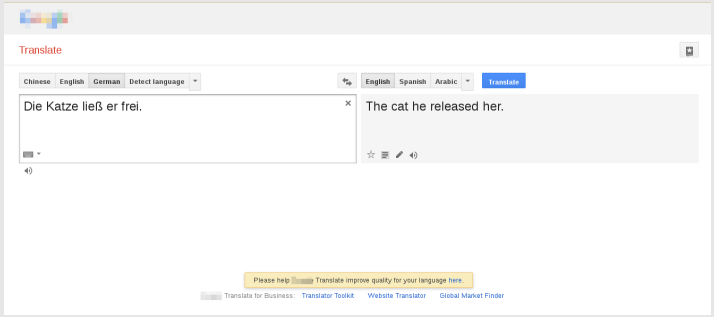
Overview

Statistical
Machine Translation



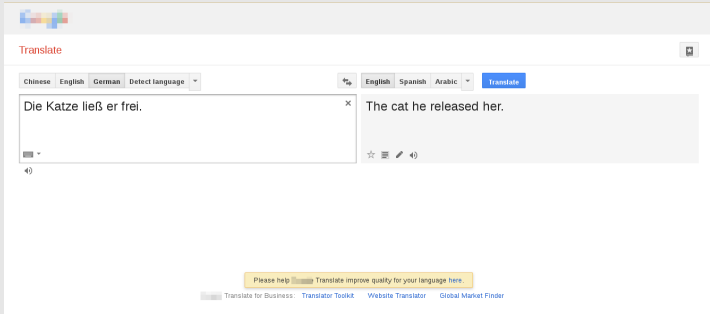
Overview

Statistical Machine Translation



Overview

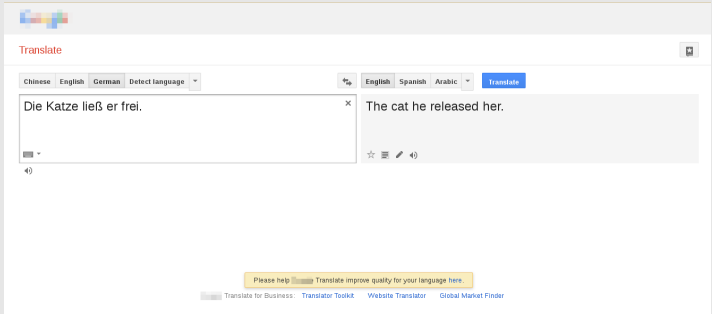
Statistical Machine Translation

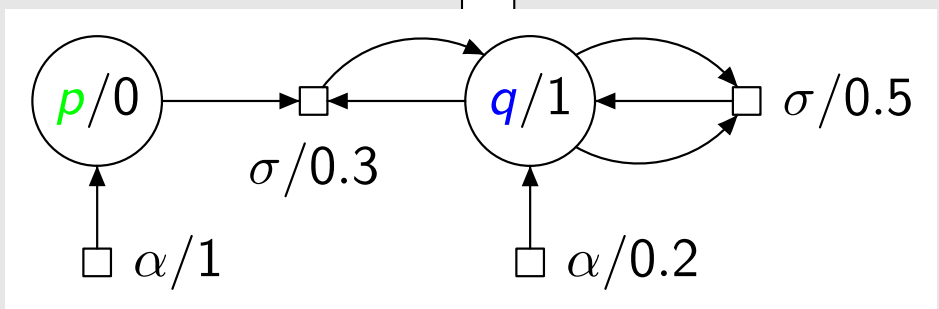


Formal-Language Theory

Overview

Statistical Machine Translation

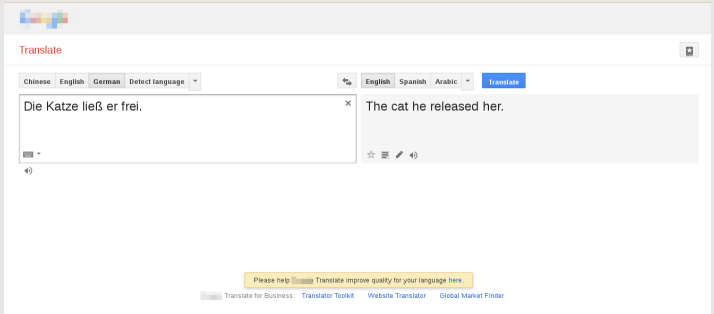




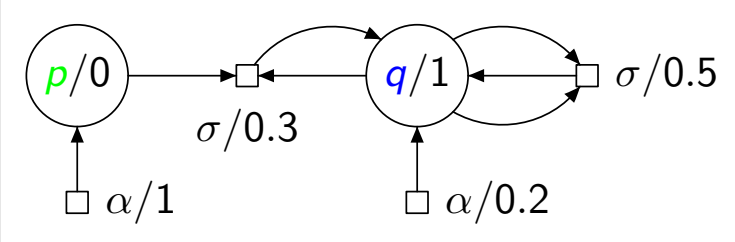
Formal-Language Theory

Overview

Statistical Machine Translation

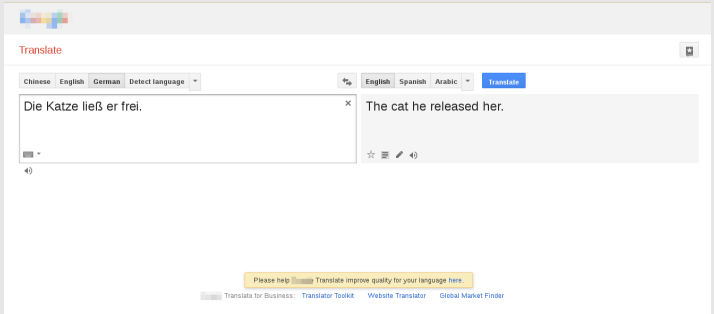


Formal-Language Theory



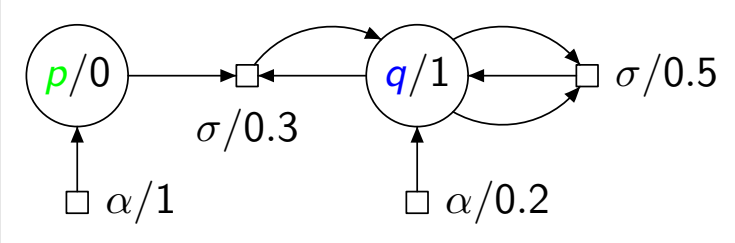
Overview

Statistical Machine Translation



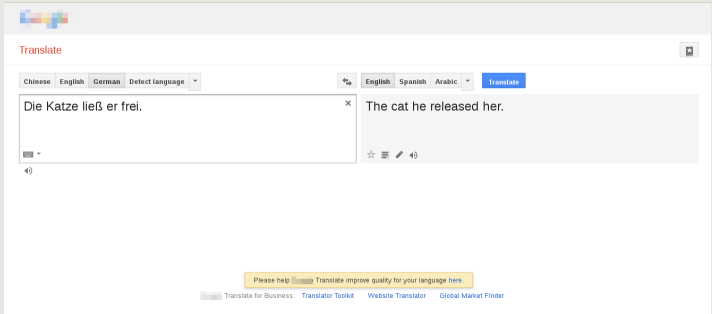
Decoder specification

Formal-Language Theory



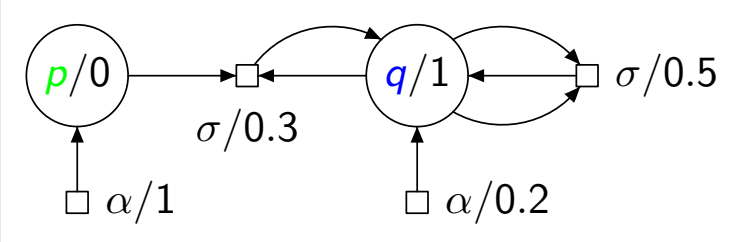
Overview

Statistical Machine Translation



Decoder specification

Formal-Language Theory



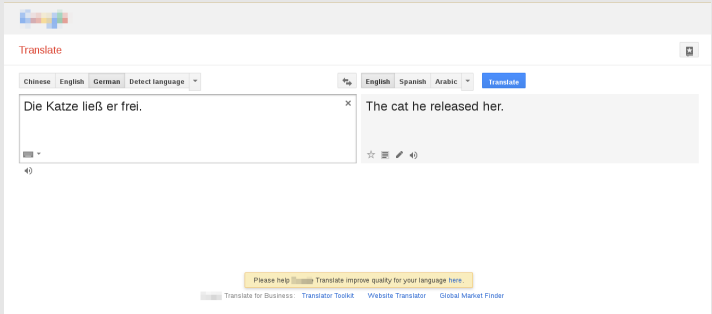
closure properties

binarization

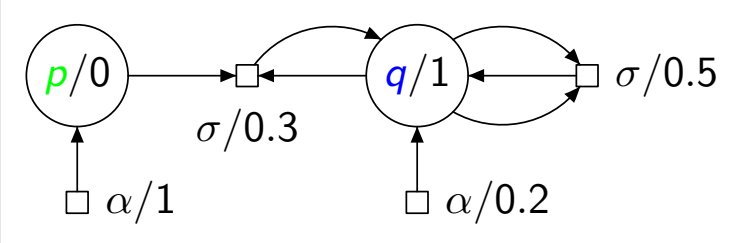
determinization

Overview

Statistical Machine Translation



Formal-Language Theory



Decoder specification

algebraic decoder specification

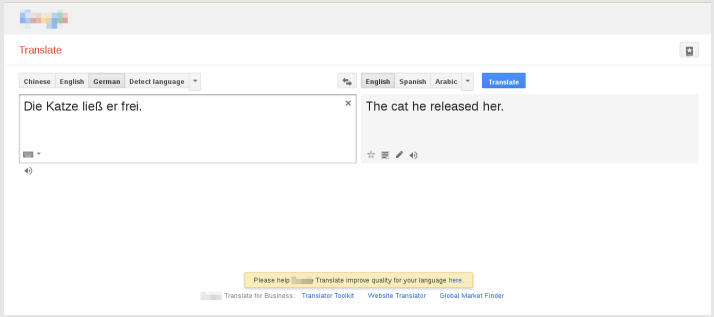
closure properties

binarization

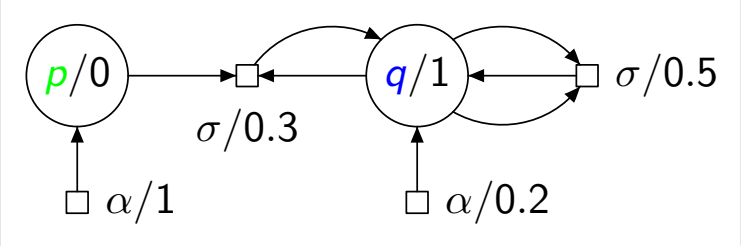
determinization

Overview

Statistical Machine Translation



Formal-Language Theory



Decoder specification

algebraic decoder specification

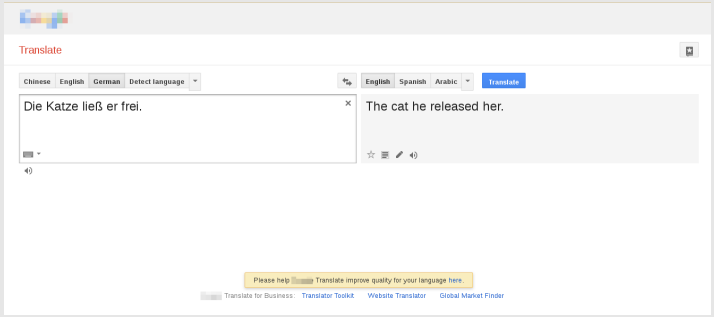
closure properties

binarization

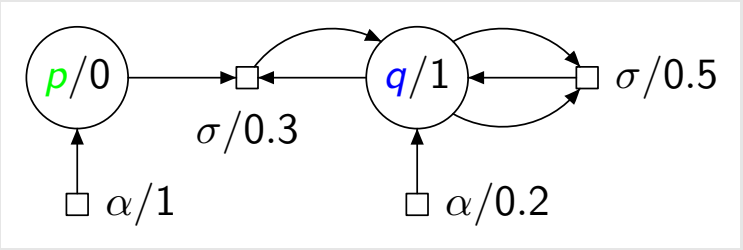
determinization

Overview

Statistical Machine Translation



Formal-Language Theory



1 2 3

Decoder specification

algebraic decoder specification

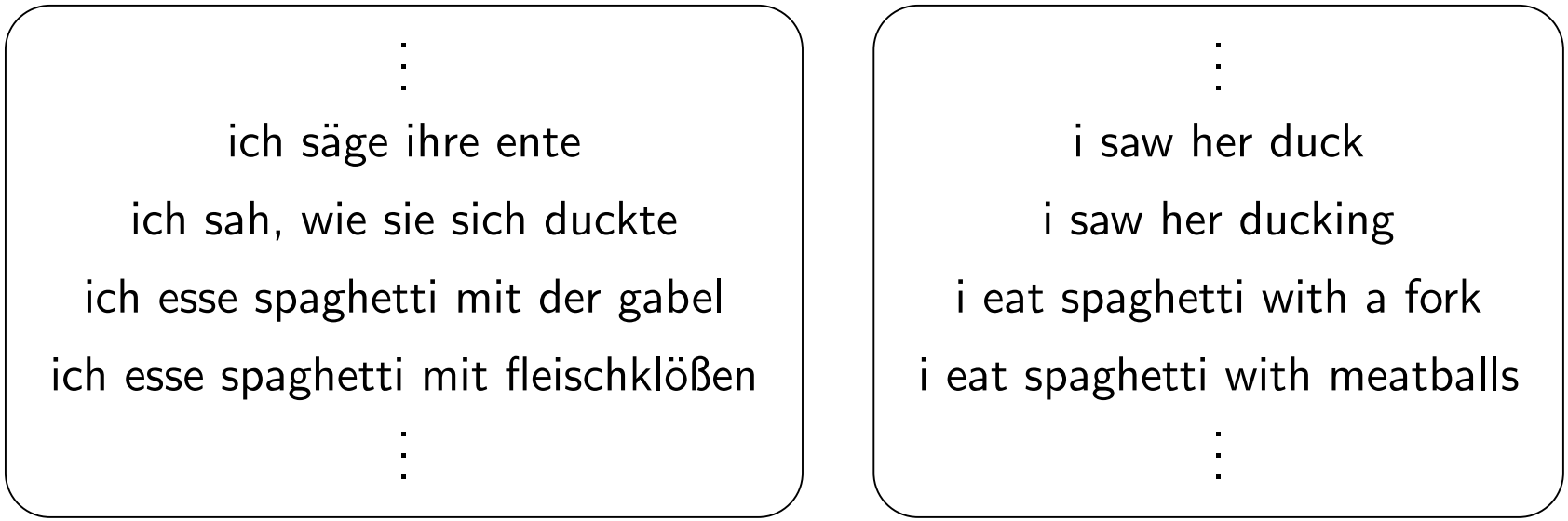
closure properties

binarization

determinization

Outline

Statistical Machine Translation



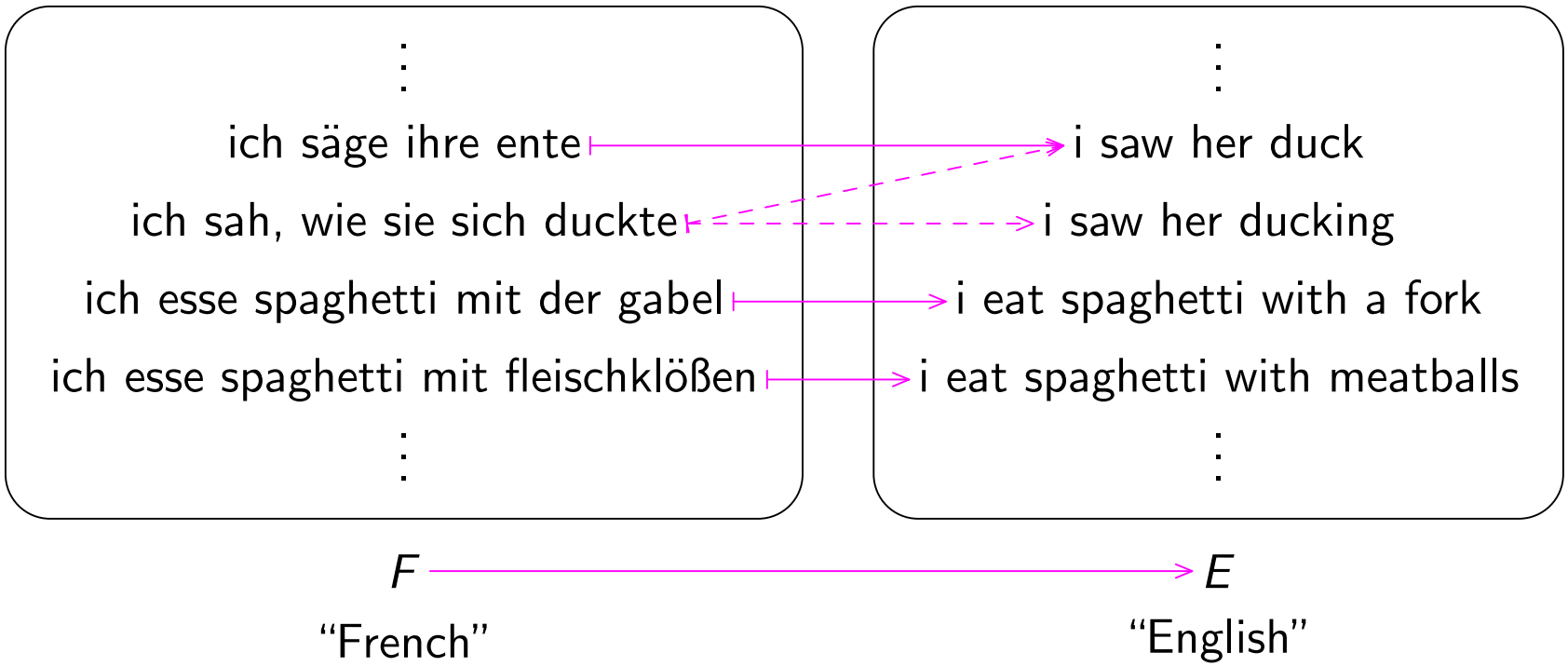
F

"French"

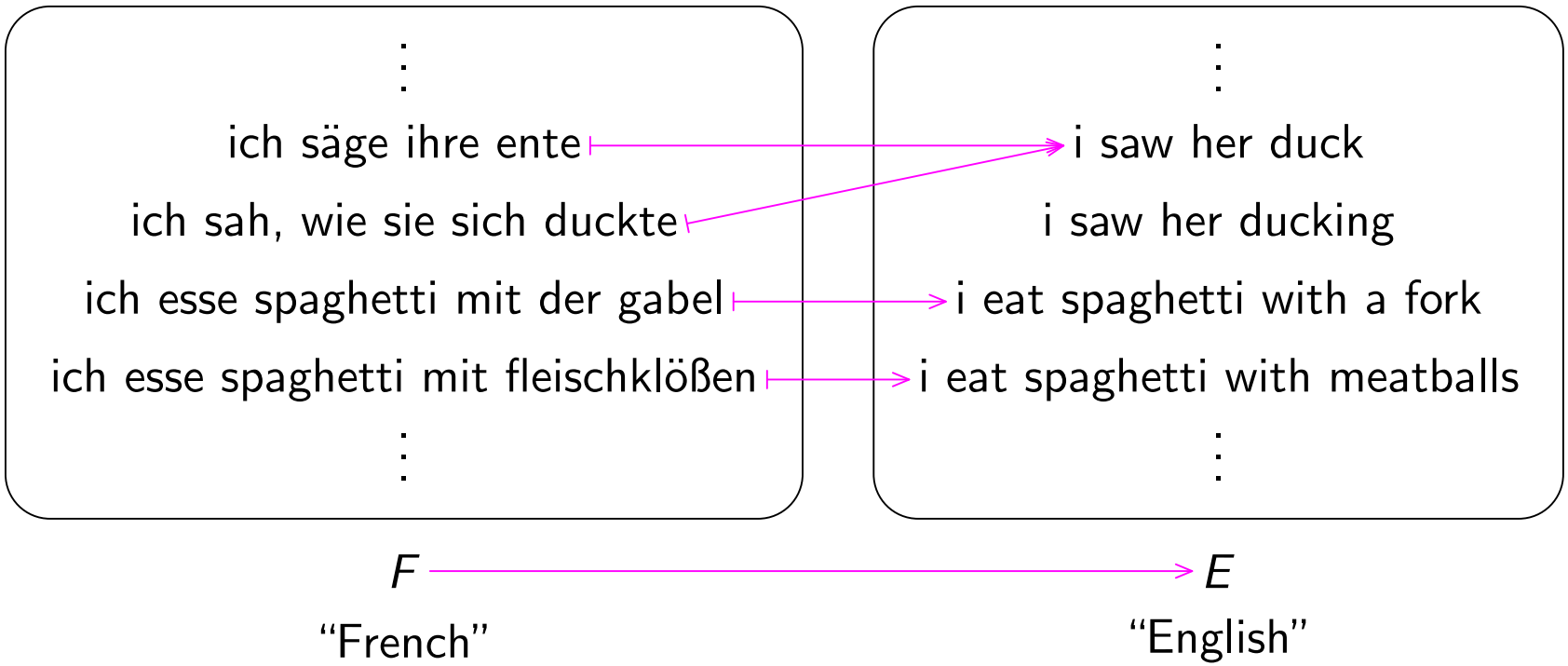
E

"English"

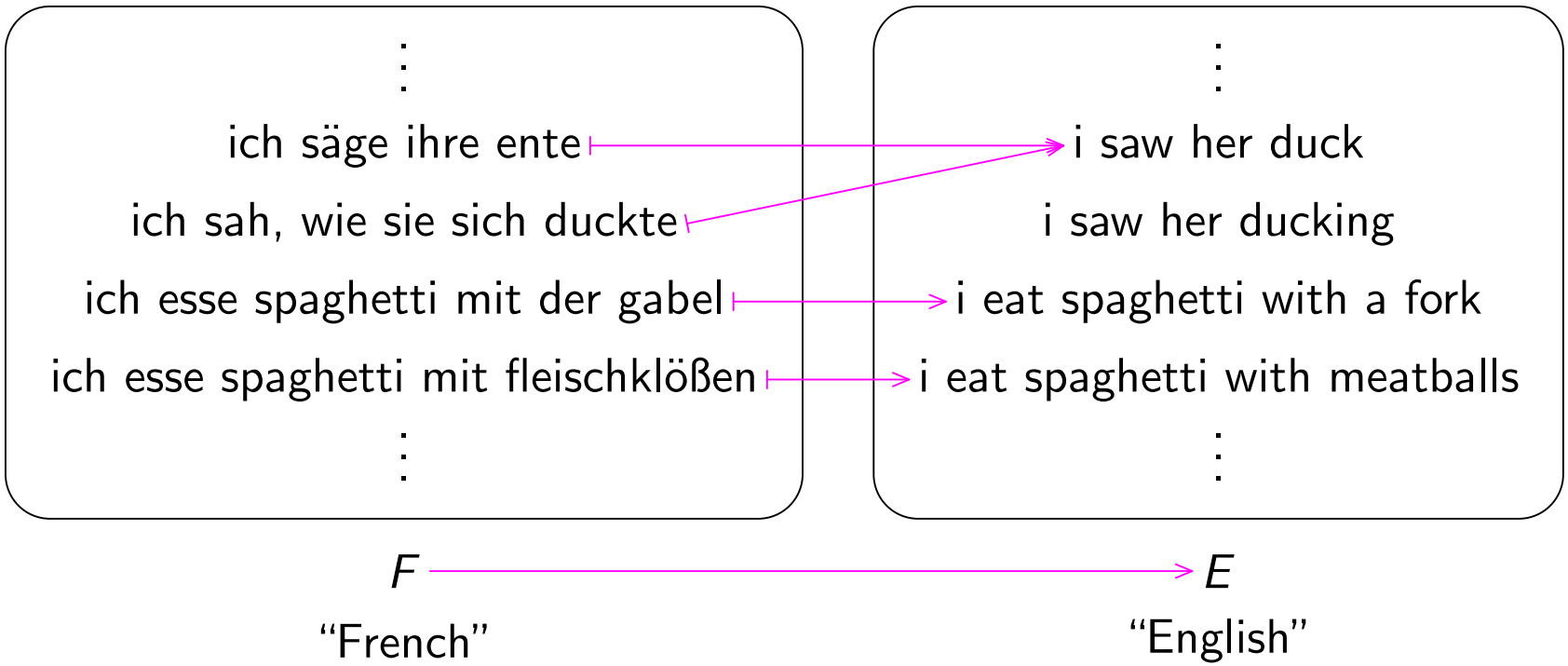
Statistical Machine Translation



Statistical Machine Translation



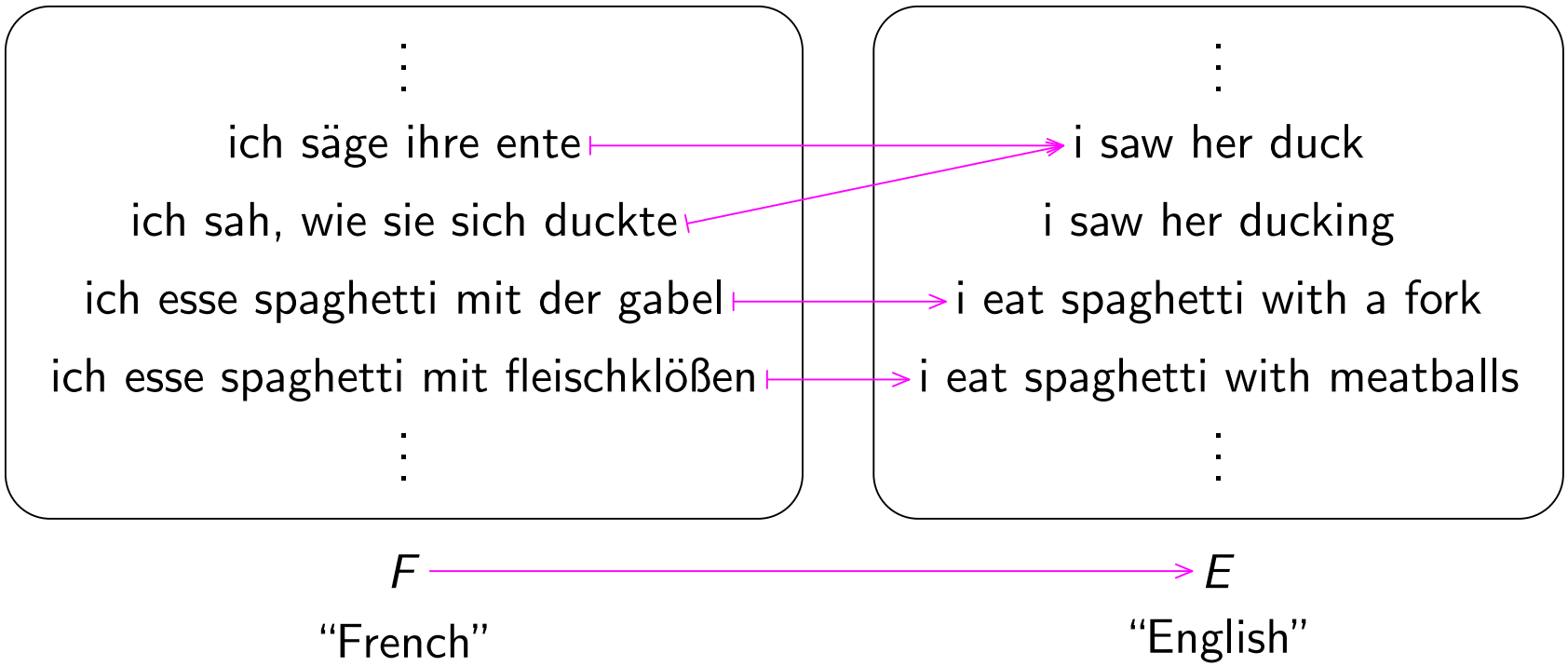
Statistical Machine Translation



Decoder

$\mathbb{D}$: $F \to E$

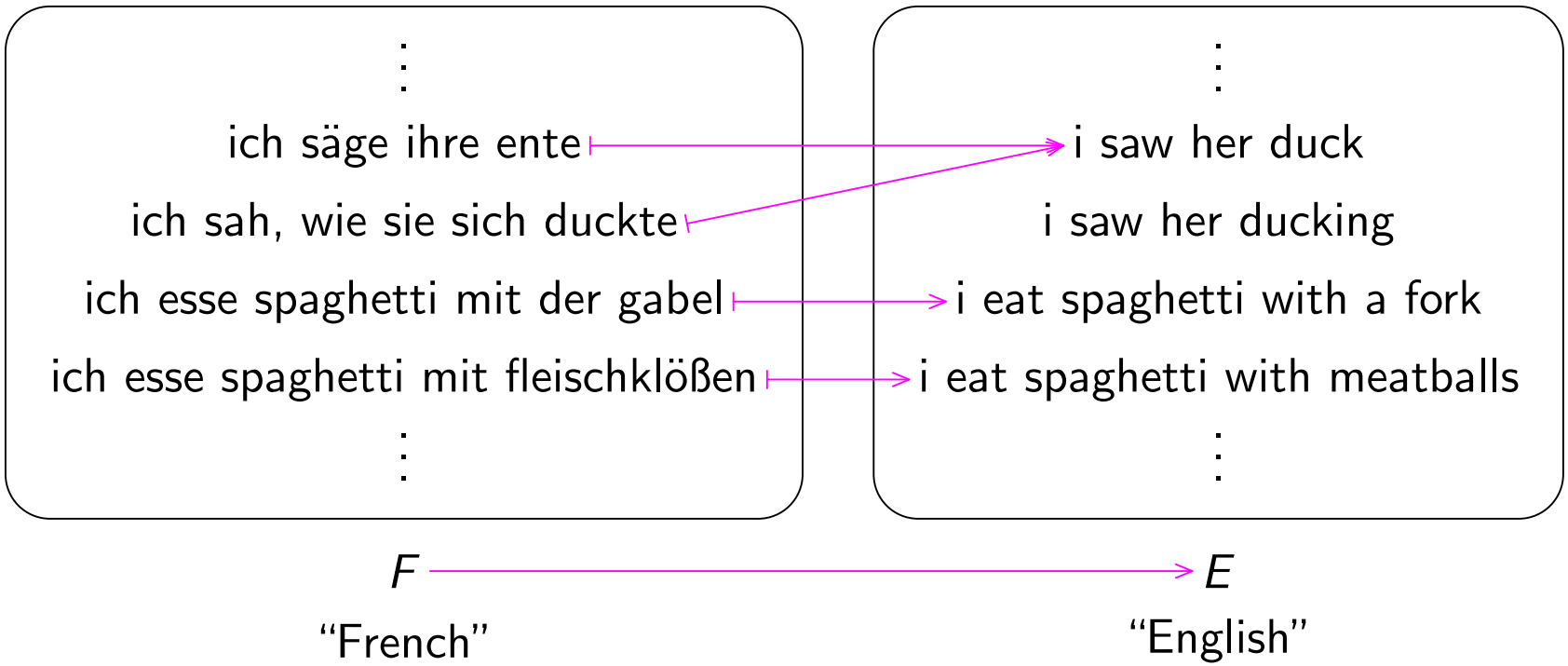
Statistical Machine Translation



Decoder

$\mathbb{D}: \Omega \to (F \to E)$ $\quad$ Ω . . . parameter space

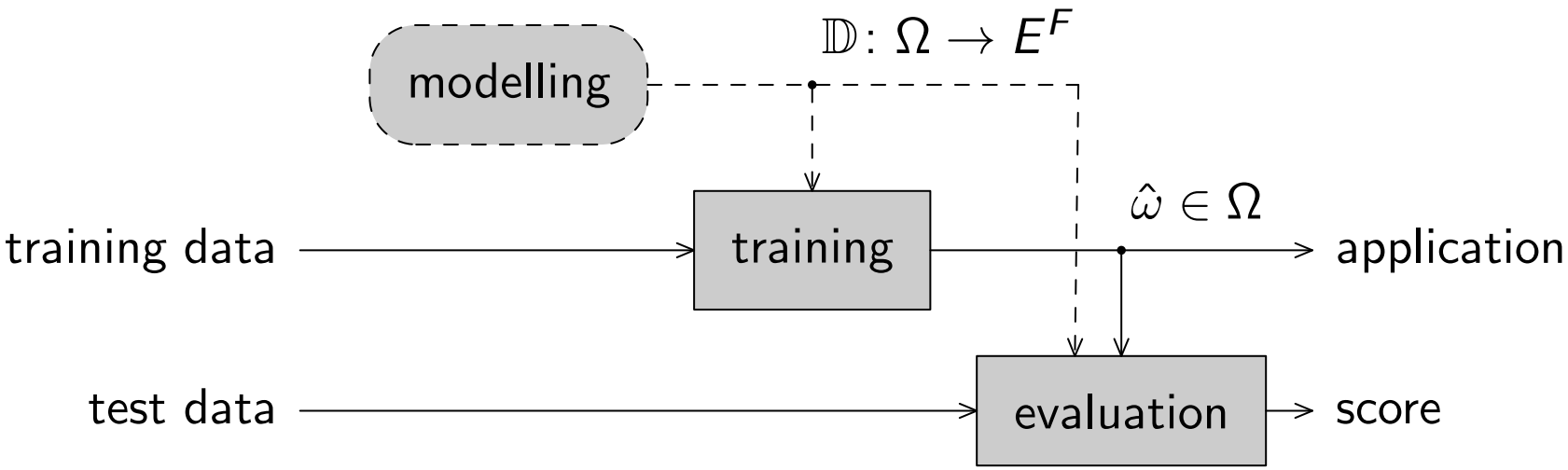
Statistical Machine Translation



Decoder

$\mathbb{D}\colon \Omega \to E^F$ $\quad\quad$ Ω . . . parameter space

Development cycle



Modelling

Hierarchical Phrases

die katze ließ er frei

$$\omega \in \Omega: \quad \begin{array}{rcl} x_1 \text{ ließ } x_2 \text{ frei} & \longleftrightarrow & x_2 \text{ freed } x_1 \\ \text{die katze} & \longleftrightarrow & \text{the cat} \\ \text{er} & \longleftrightarrow & \text{he} \end{array}$$

Modelling

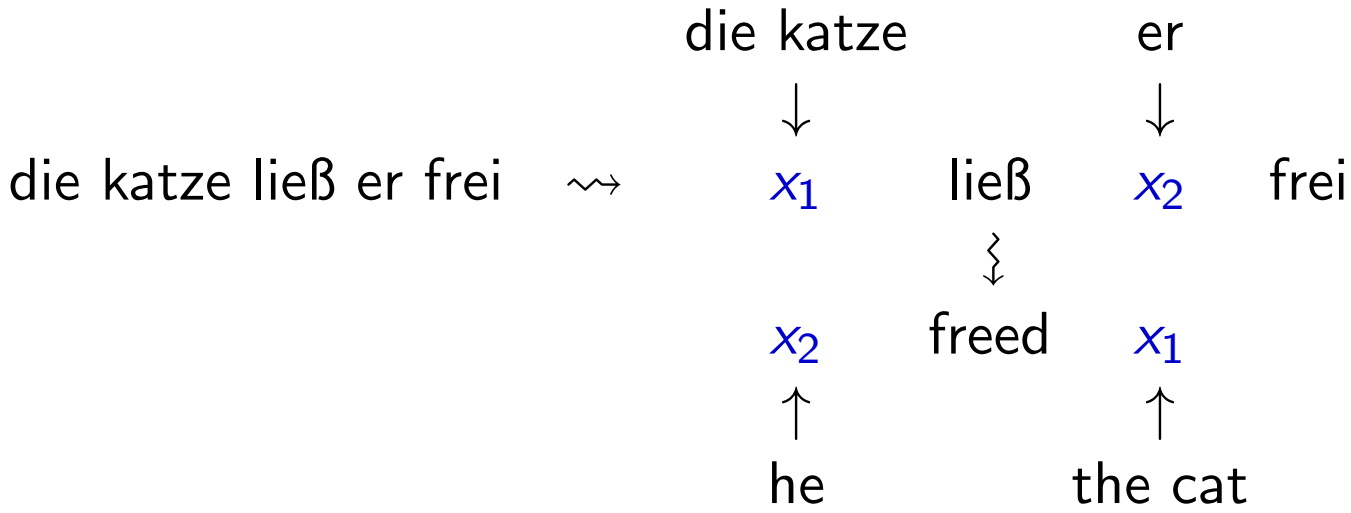
Hierarchical Phrases

$$\begin{array}{lcccccc} & & \text{die katze} & & \text{er} & & \\ & & \downarrow & & \downarrow & & \\ \text{die katze ließ er frei} & \rightsquigarrow & x_1 & \text{ließ} & x_2 & \text{frei} & \end{array}$$

$$\begin{array}{rrcl} & x_1 \text{ ließ } x_2 \text{ frei} & \longleftrightarrow & x_2 \text{ freed } x_1 \\ \omega \in \Omega: & \text{die katze} & \longleftrightarrow & \text{the cat} \\ & \text{er} & \longleftrightarrow & \text{he} \end{array}$$

Modelling

Hierarchical Phrases

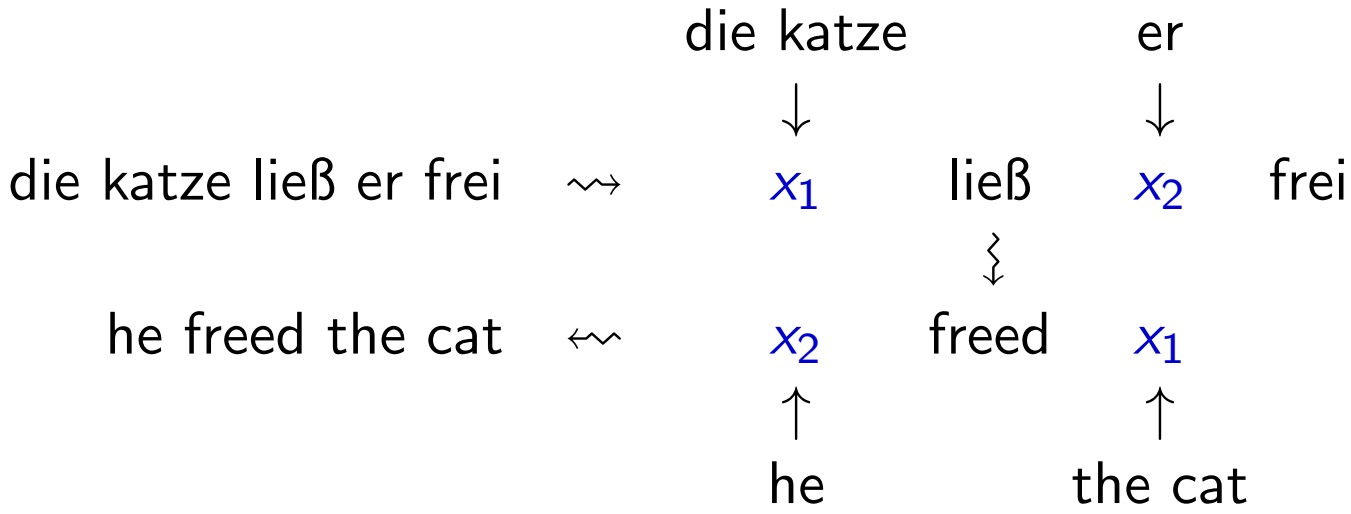


$\omega \in \Omega$:

$$x_1 \text{ ließ } x_2 \text{ frei} \longleftrightarrow x_2 \text{ freed } x_1$$
$$\text{die katze} \longleftrightarrow \text{the cat}$$
$$\text{er} \longleftrightarrow \text{he}$$

Modelling

Hierarchical Phrases



$\omega \in \Omega$:

$$x_1 \text{ ließ } x_2 \text{ frei} \longleftrightarrow x_2 \text{ freed } x_1$$
$$\text{die katze} \longleftrightarrow \text{the cat}$$
$$\text{er} \longleftrightarrow \text{he}$$

Training

parallel corpus

$d \in (E \times F)^*$

Resumption of the session	Wiederaufnahme der Sitzungsperiode
I declare resumed the session of the European Parliament adjourned on Friday 17 December 1999 , [. . .] .	Ich erkläre die am Freitag , dem 17. Dezember unterbrochene Sitzungsperiode des Europäischen Parlaments für wiederaufgenommen , [. . .] .
⋮	⋮

EuroParl corpus, 11 languages, 1.5 million sentences each

Training

parallel corpus

$d \in (E \times F)^*$

$\downarrow$

apply heuristics and statistical methods

Training

parallel corpus

$d \in (E \times F)^*$

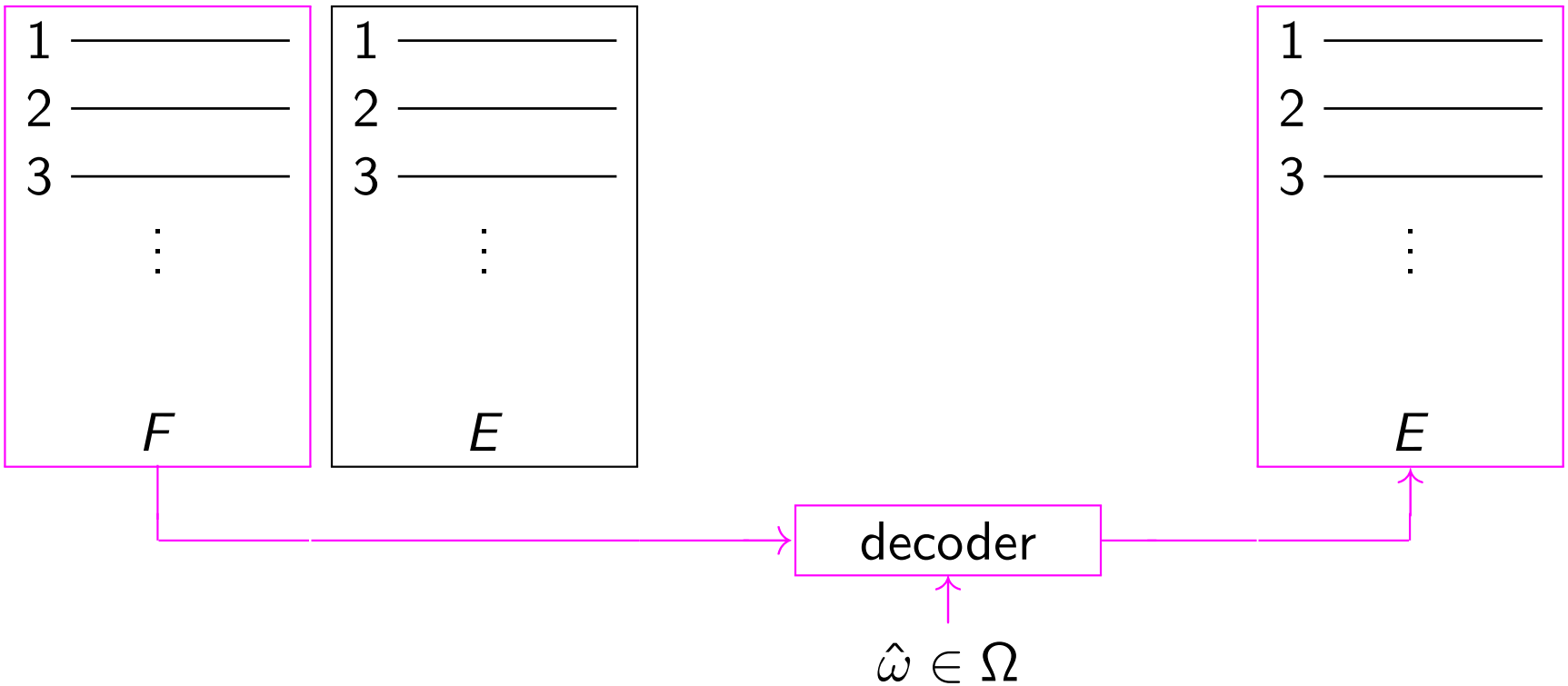
↓

apply heuristics and statistical methods

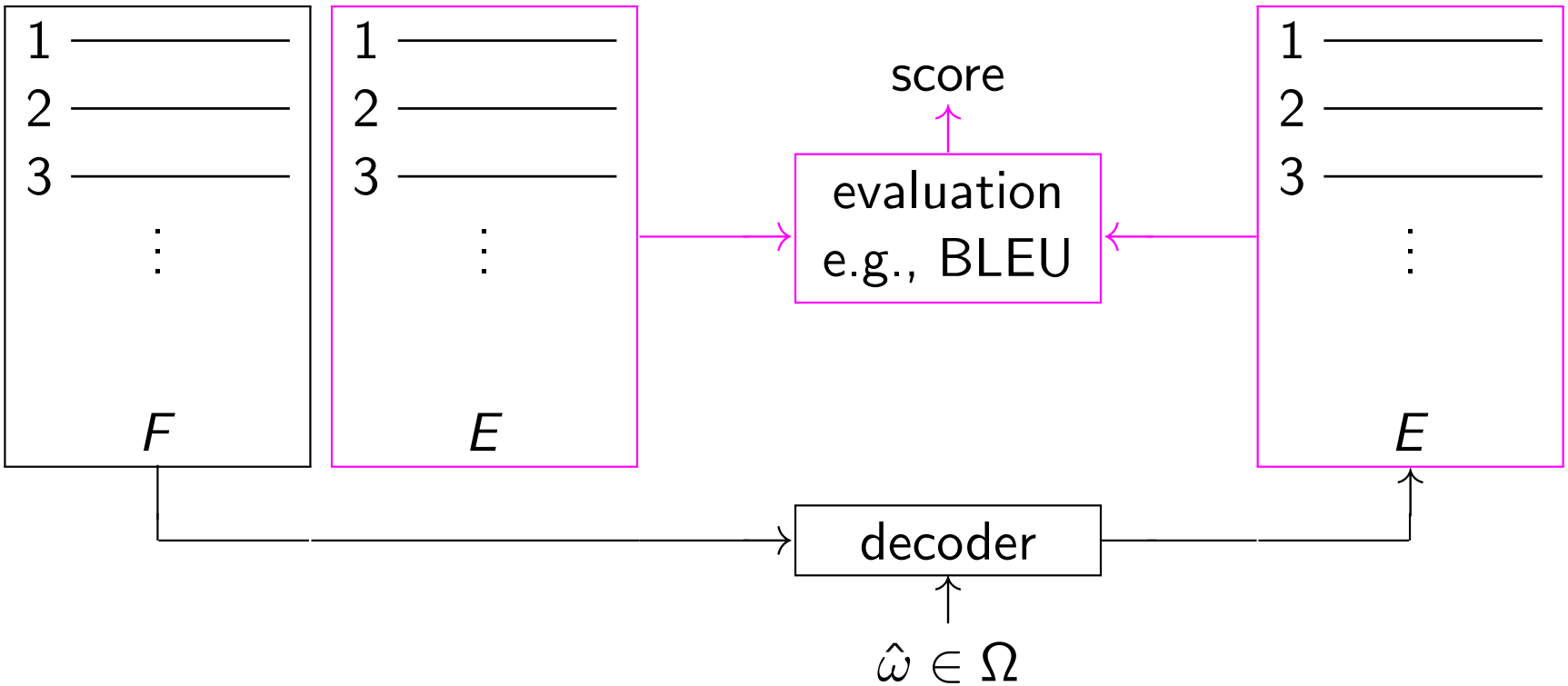
↓

$\hat{\omega} \in \Omega$

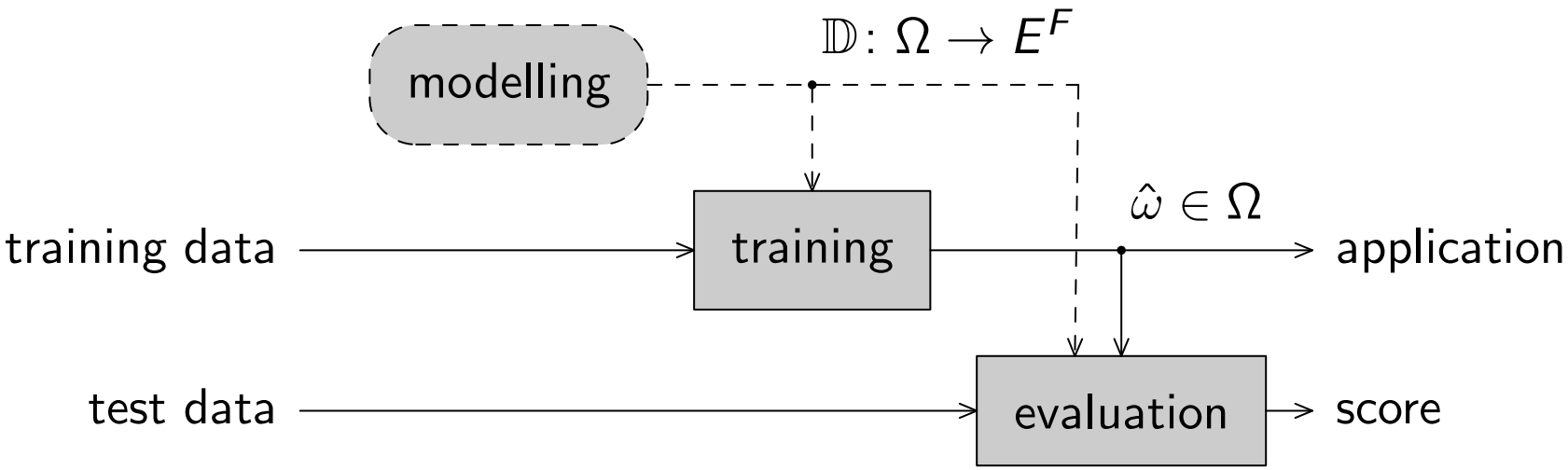
Evaluation



Evaluation


1
2
3
F
1
2
3
E
score
evaluation
e.g., BLEU
1
2
3
E
decoder
$\hat{\omega} \in \Omega$

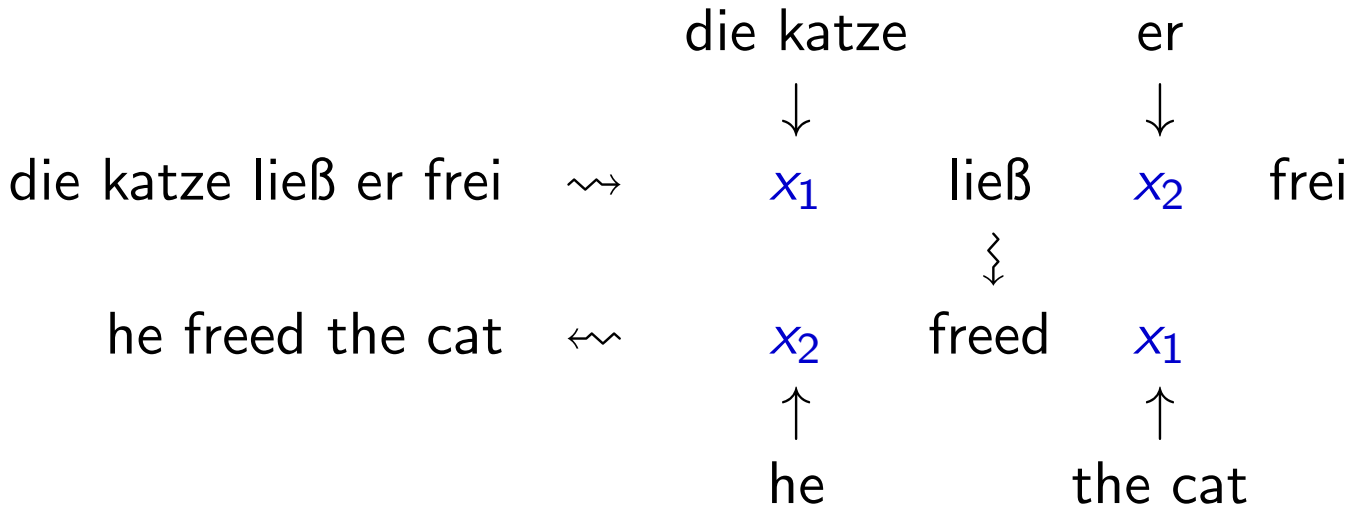
Development cycle



Outline

Idea

Hierarchical Phrases



$\omega \in \Omega$:

$$x_1 \text{ ließ } x_2 \text{ frei} \longleftrightarrow x_2 \text{ freed } x_1$$
$$\text{die katze} \longleftrightarrow \text{the cat}$$
$$\text{er} \longleftrightarrow \text{he}$$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

ρ_1: $S \rightarrow \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle (NP)$

ρ_2: $S \rightarrow \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle (PPER)$

ρ_3: $S \rightarrow \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle (PPER, NP)$

ρ_4: $S \rightarrow \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle (PPER, NP)$

ρ_5: $PPER \rightarrow \langle \text{er}, \text{he} \rangle$

ρ_6: $NP \rightarrow \langle \text{die katze}, \text{the cat} \rangle$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

ρ_1: $S \rightarrow \alpha_1(NP)$ $\quad \alpha_1 = \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle$

ρ_2: $S \rightarrow \alpha_2(PPER)$ $\quad \alpha_2 = \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle$

ρ_3: $S \rightarrow \alpha_3(PPER, NP)$ $\quad \alpha_3 = \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_4: $S \rightarrow \alpha_4(PPER, NP)$ $\quad \alpha_4 = \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_5: $PPER \rightarrow \alpha_5$ $\quad \alpha_5 = \langle \text{er}, \text{he} \rangle$

ρ_6: $NP \rightarrow \alpha_6$ $\quad \alpha_6 = \langle \text{die katze}, \text{the cat} \rangle$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

ρ_1: $S \rightarrow \alpha_1(NP)$ $\quad \alpha_1 = \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle$

ρ_2: $S \rightarrow \alpha_2(PPER)$ $\quad \alpha_2 = \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle$

ρ_3: $S \rightarrow \alpha_3(PPER, NP)$ $\quad \alpha_3 = \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_4: $S \rightarrow \alpha_4(PPER, NP)$ $\quad \alpha_4 = \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_5: $PPER \rightarrow \alpha_5$ $\quad \alpha_5 = \langle \text{er}, \text{he} \rangle$

ρ_6: $NP \rightarrow \alpha_6$ $\quad \alpha_6 = \langle \text{die katze}, \text{the cat} \rangle$

S

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

ρ_1: $S \rightarrow \alpha_1(NP)$ $\quad \alpha_1 = \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle$

ρ_2: $S \rightarrow \alpha_2(PPER)$ $\quad \alpha_2 = \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle$

ρ_3: $S \rightarrow \alpha_3(PPER, NP)$ $\quad \alpha_3 = \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_4: $S \rightarrow \alpha_4(PPER, NP)$ $\quad \alpha_4 = \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_5: $PPER \rightarrow \alpha_5$ $\quad \alpha_5 = \langle \text{er}, \text{he} \rangle$

ρ_6: $NP \rightarrow \alpha_6$ $\quad \alpha_6 = \langle \text{die katze}, \text{the cat} \rangle$

$$S \stackrel{\rho_4}{\Longrightarrow} \alpha_4(PPER, NP)$$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

$$
\begin{array}{lrll}
\rho_1\colon & S & \to \alpha_1(NP) & \alpha_1 = \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle \\
\rho_2\colon & S & \to \alpha_2(PPER) & \alpha_2 = \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle \\
\rho_3\colon & S & \to \alpha_3(PPER, NP) & \alpha_3 = \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle \\
\rho_4\colon & S & \to \alpha_4(PPER, NP) & \alpha_4 = \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle \\
\rho_5\colon & PPER & \to \alpha_5 & \alpha_5 = \langle \text{er}, \text{he} \rangle \\
\rho_6\colon & NP & \to \alpha_6 & \alpha_6 = \langle \text{die katze}, \text{the cat} \rangle
\end{array}
$$

$$
S \overset{\rho_4}{\Longrightarrow} \alpha_4(PPER, NP) \overset{\rho_5}{\Longrightarrow} \alpha_4(\alpha_5, NP)
$$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

$$
\begin{array}{lrll}
\rho_1\colon & S & \to \alpha_1(NP) & \alpha_1 = \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle \\
\rho_2\colon & S & \to \alpha_2(PPER) & \alpha_2 = \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle \\
\rho_3\colon & S & \to \alpha_3(PPER, NP) & \alpha_3 = \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle \\
\rho_4\colon & S & \to \alpha_4(PPER, NP) & \alpha_4 = \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle \\
\rho_5\colon & PPER & \to \alpha_5 & \alpha_5 = \langle \text{er}, \text{he} \rangle \\
\rho_6\colon & NP & \to \alpha_6 & \alpha_6 = \langle \text{die katze}, \text{the cat} \rangle
\end{array}
$$

$$
S \overset{\rho_4}{\Longrightarrow} \alpha_4(PPER, NP) \overset{\rho_5}{\Longrightarrow} \alpha_4(\alpha_5, NP) \overset{\rho_6}{\Longrightarrow} \alpha_4(\alpha_5, \alpha_6)
$$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

$$
\begin{array}{lrll}
\rho_1\colon & S & \to \alpha_1(NP) & \alpha_1 = \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle \\
\rho_2\colon & S & \to \alpha_2(PPER) & \alpha_2 = \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle \\
\rho_3\colon & S & \to \alpha_3(PPER, NP) & \alpha_3 = \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle \\
\rho_4\colon & S & \to \alpha_4(PPER, NP) & \alpha_4 = \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle \\
\rho_5\colon & PPER & \to \alpha_5 & \alpha_5 = \langle \text{er}, \text{he} \rangle \\
\rho_6\colon & NP & \to \alpha_6 & \alpha_6 = \langle \text{die katze}, \text{the cat} \rangle
\end{array}
$$

$$
S \overset{\rho_4}{\Longrightarrow} \alpha_4(PPER, NP) \overset{\rho_5}{\Longrightarrow} \alpha_4(\alpha_5, NP) \overset{\rho_6}{\Longrightarrow} \alpha_4(\alpha_5, \alpha_6) \qquad \rho_4(\rho_5, \rho_6)
$$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

ρ_1: $S \to \alpha_1(NP)$ $\quad \alpha_1 = \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle$

ρ_2: $S \to \alpha_2(PPER)$ $\quad \alpha_2 = \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle$

ρ_3: $S \to \alpha_3(PPER, NP)$ $\quad \alpha_3 = \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_4: $S \to \alpha_4(PPER, NP)$ $\quad \alpha_4 = \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_5: $PPER \to \alpha_5$ $\quad \alpha_5 = \langle \text{er}, \text{he} \rangle$

ρ_6: $NP \to \alpha_6$ $\quad \alpha_6 = \langle \text{die katze}, \text{the cat} \rangle$

$$S \overset{\rho_4}{\Longrightarrow} \alpha_4(PPER, NP) \overset{\rho_5}{\Longrightarrow} \alpha_4(\alpha_5, NP) \overset{\rho_6}{\Longrightarrow} \alpha_4(\alpha_5, \alpha_6) \qquad \rho_4(\rho_5, \rho_6)$$

$h_1(\alpha_4(\alpha_5, \alpha_6))$

$h_2(\alpha_4(\alpha_5, \alpha_6))$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

ρ_1: $S \to \alpha_1(NP)$ $\quad \alpha_1 = \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle$

ρ_2: $S \to \alpha_2(PPER)$ $\quad \alpha_2 = \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle$

ρ_3: $S \to \alpha_3(PPER, NP)$ $\quad \alpha_3 = \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_4: $S \to \alpha_4(PPER, NP)$ $\quad \alpha_4 = \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_5: $PPER \to \alpha_5$ $\quad \alpha_5 = \langle \text{er}, \text{he} \rangle$

ρ_6: $NP \to \alpha_6$ $\quad \alpha_6 = \langle \text{die katze}, \text{the cat} \rangle$

$$S \overset{\rho_4}{\Longrightarrow} \alpha_4(PPER, NP) \overset{\rho_5}{\Longrightarrow} \alpha_4(\alpha_5, NP) \overset{\rho_6}{\Longrightarrow} \alpha_4(\alpha_5, \alpha_6) \qquad \rho_4(\rho_5, \rho_6)$$

$$h_1(\alpha_4(\alpha_5, \alpha_6)) = h_1(\alpha_6) \text{ ließ } h_1(\alpha_5) \text{ frei}$$

$$h_2(\alpha_4(\alpha_5, \alpha_6)) = h_2(\alpha_5) \text{ freed } h_2(\alpha_6)$$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

ρ_1: $S \to \alpha_1(NP)$ $\quad \alpha_1 = \langle x_1$ ließ er frei, he freed $x_1 \rangle$

ρ_2: $S \to \alpha_2(PPER)$ $\quad \alpha_2 = \langle$die katze ließ x_1 frei, x_1 let the cat out$\rangle$

ρ_3: $S \to \alpha_3(PPER, NP)$ $\quad \alpha_3 = \langle x_1$ ließ x_2 frei, x_1 freed $x_2 \rangle$

ρ_4: $S \to \alpha_4(PPER, NP)$ $\quad \alpha_4 = \langle x_2$ ließ x_1 frei, x_1 freed $x_2 \rangle$

ρ_5: $PPER \to \alpha_5$ $\quad \alpha_5 = \langle$er, he$\rangle$

ρ_6: $NP \to \alpha_6$ $\quad \alpha_6 = \langle$die katze, the cat$\rangle$

$$S \overset{\rho_4}{\Longrightarrow} \alpha_4(PPER, NP) \overset{\rho_5}{\Longrightarrow} \alpha_4(\alpha_5, NP) \overset{\rho_6}{\Longrightarrow} \alpha_4(\alpha_5, \alpha_6) \qquad \rho_4(\rho_5, \rho_6)$$

$$h_1(\alpha_4(\alpha_5, \alpha_6)) = h_1(\alpha_6) \text{ ließ } h_1(\alpha_5) \text{ frei} = \text{die katze ließ er frei}$$

$$h_2(\alpha_4(\alpha_5, \alpha_6)) = h_2(\alpha_5) \text{ freed } h_2(\alpha_6) \quad = \text{he freed the cat}$$

Synchronous Context-Free Grammars (SCFGs)

(Lewis und Stearns 1966)

ρ_1: $S \to \alpha_1(NP)$ $\quad \alpha_1 = \langle x_1 \text{ ließ er frei}, \text{he freed } x_1 \rangle$

ρ_2: $S \to \alpha_2(PPER)$ $\quad \alpha_2 = \langle \text{die katze ließ } x_1 \text{ frei}, x_1 \text{ let the cat out} \rangle$

ρ_3: $S \to \alpha_3(PPER, NP)$ $\quad \alpha_3 = \langle x_1 \text{ ließ } x_2 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_4: $S \to \alpha_4(PPER, NP)$ $\quad \alpha_4 = \langle x_2 \text{ ließ } x_1 \text{ frei}, x_1 \text{ freed } x_2 \rangle$

ρ_5: $PPER \to \alpha_5$ $\quad \alpha_5 = \langle \text{er}, \text{he} \rangle$

ρ_6: $NP \to \alpha_6$ $\quad \alpha_6 = \langle \text{die katze}, \text{the cat} \rangle$

$$S \overset{\rho_4}{\Longrightarrow} \alpha_4(PPER, NP) \overset{\rho_5}{\Longrightarrow} \alpha_4(\alpha_5, NP) \overset{\rho_6}{\Longrightarrow} \alpha_4(\alpha_5, \alpha_6) \qquad \rho_4(\rho_5, \rho_6)$$

$$h_1(\alpha_4(\alpha_5, \alpha_6)) = h_1(\alpha_6) \text{ ließ } h_1(\alpha_5) \text{ frei} = \text{die katze ließ er frei}$$
$$h_2(\alpha_4(\alpha_5, \alpha_6)) = h_2(\alpha_5) \text{ freed } h_2(\alpha_6) \quad = \text{he freed the cat}$$

Abstract Syntax Trees

$D^S(G) = \{\rho_1(\rho_6), \rho_2(\rho_5), \rho_3(\rho_5, \rho_6), \rho_4(\rho_5, \rho_6)\}$

$D^{PPER}(G) = \{\rho_5\}$, $D^{NP}(G) = \{\rho_6\}$

Abstract Specification

(Chiang 2007)

$$\Omega = \{\ G \qquad \mid G \text{ SCFG} \qquad \}$$

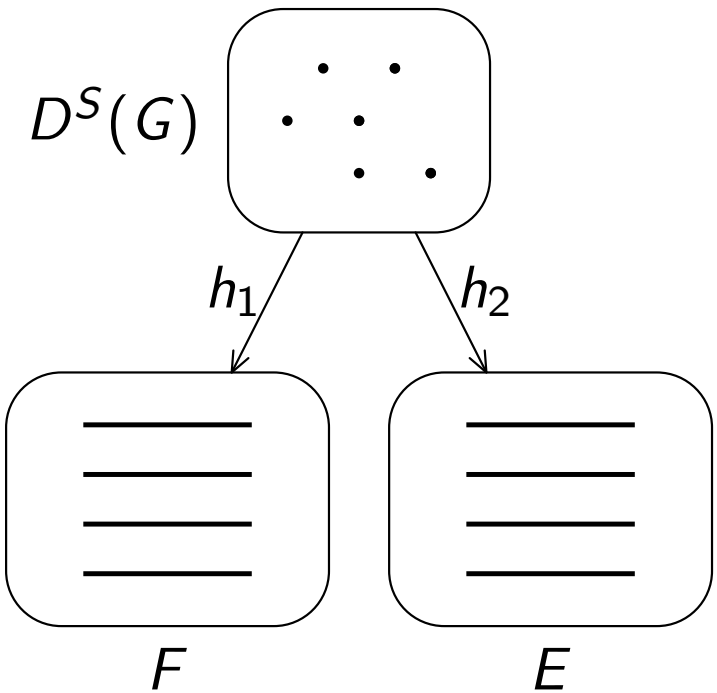
$$\mathbb{D}(G \qquad) : f \mapsto$$

Abstract Specification

(Chiang 2007)

$\Omega = \{\ G \qquad \mid G \text{ SCFG} \qquad \}$

$\mathbb{D}(G \qquad) : f \mapsto$

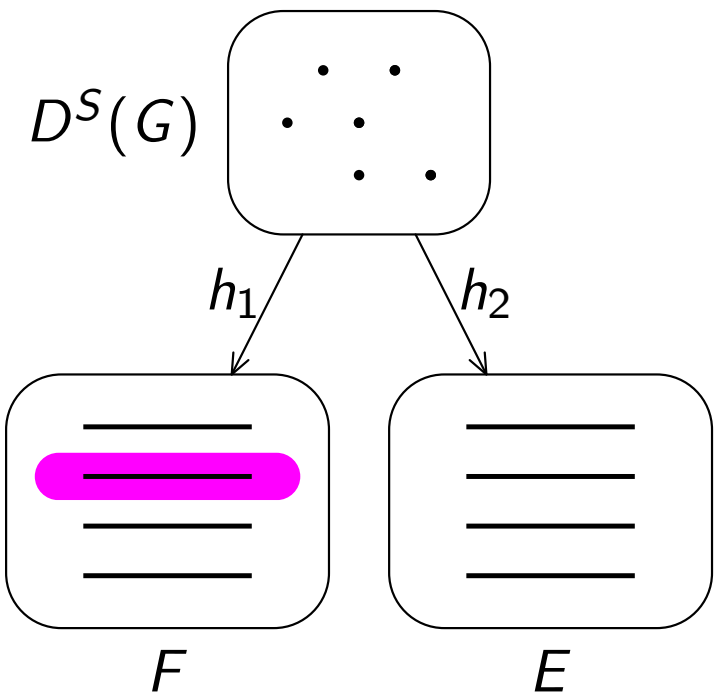


Abstract Specification

(Chiang 2007)

$\Omega = \{\ G \qquad \mid G \text{ SCFG} \qquad \}$

$\mathbb{D}(G \qquad) : f \mapsto$

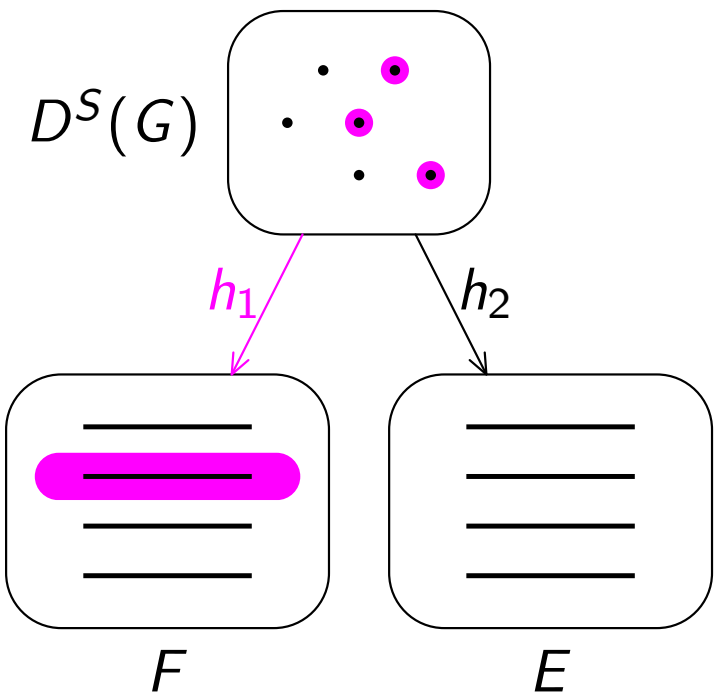


Abstract Specification

(Chiang 2007)

$\Omega = \{\ G \qquad \mid G \text{ SCFG} \qquad \}$

$\mathbb{D}(G \qquad) : f \mapsto$

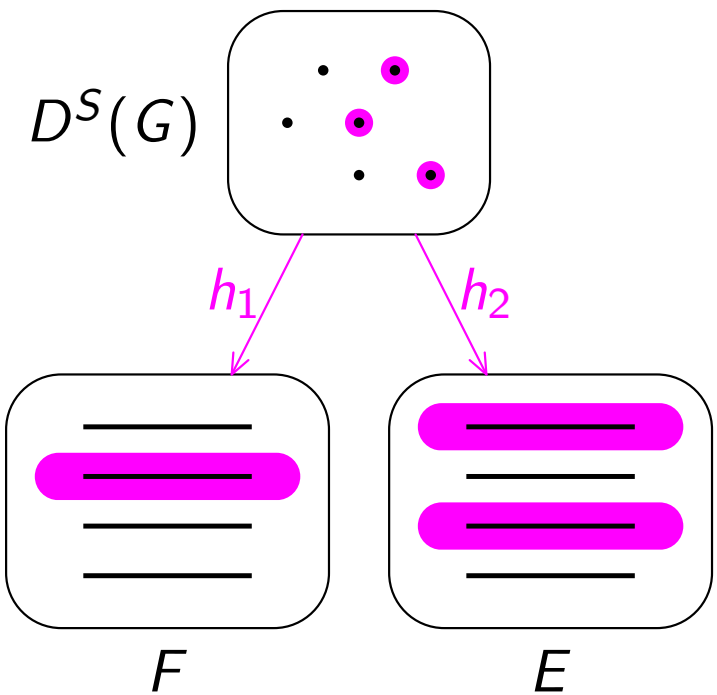


Abstract Specification

(Chiang 2007)

$\Omega = \{\ G \qquad \mid G \text{ SCFG} \qquad \}$

$\mathbb{D}(G \qquad) : f \mapsto$

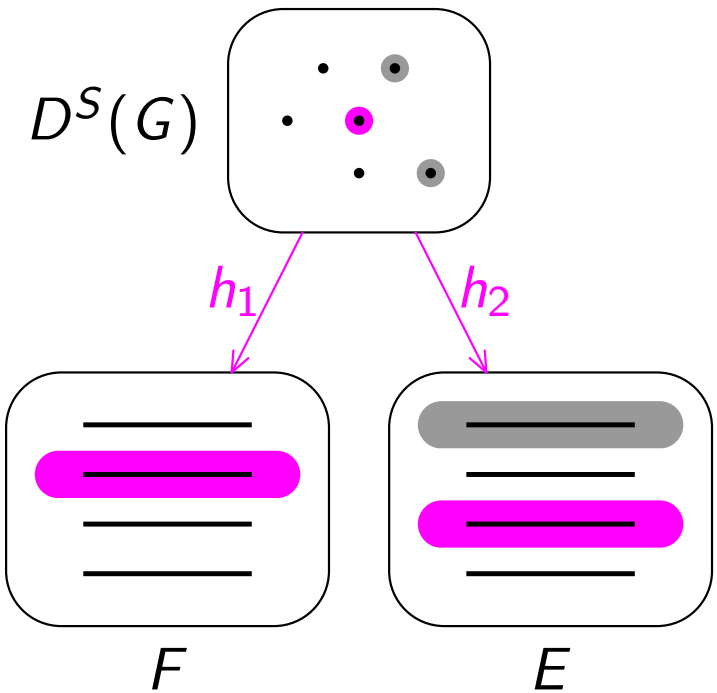


Abstract Specification

(Chiang 2007)

$$\Omega = \{\ G \quad \mid G \text{ SCFG} \quad \}$$

$$\mathbb{D}(G\ ) : f \mapsto h_2(\text{select}\ (\{d \mid d \in D^S(G), h_1(d) = f\}))$$

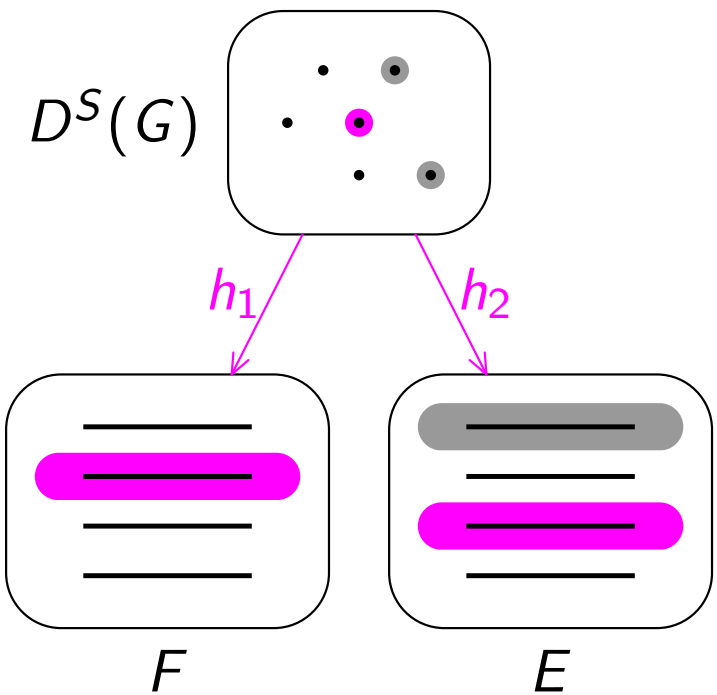


Abstract Specification

(Chiang 2007)

$$\Omega = \{(G, \mu, \theta) \mid G \text{ SCFG}, \mu \text{ probability assignment}, \theta \in \mathbb{R}^3\}$$

$$\mathbb{D}(G, \mu, \theta)\colon f \mapsto h_2(\text{select}_{G,\mu,\theta}(\{d \mid d \in D^S(G), h_1(d) = f\}))$$

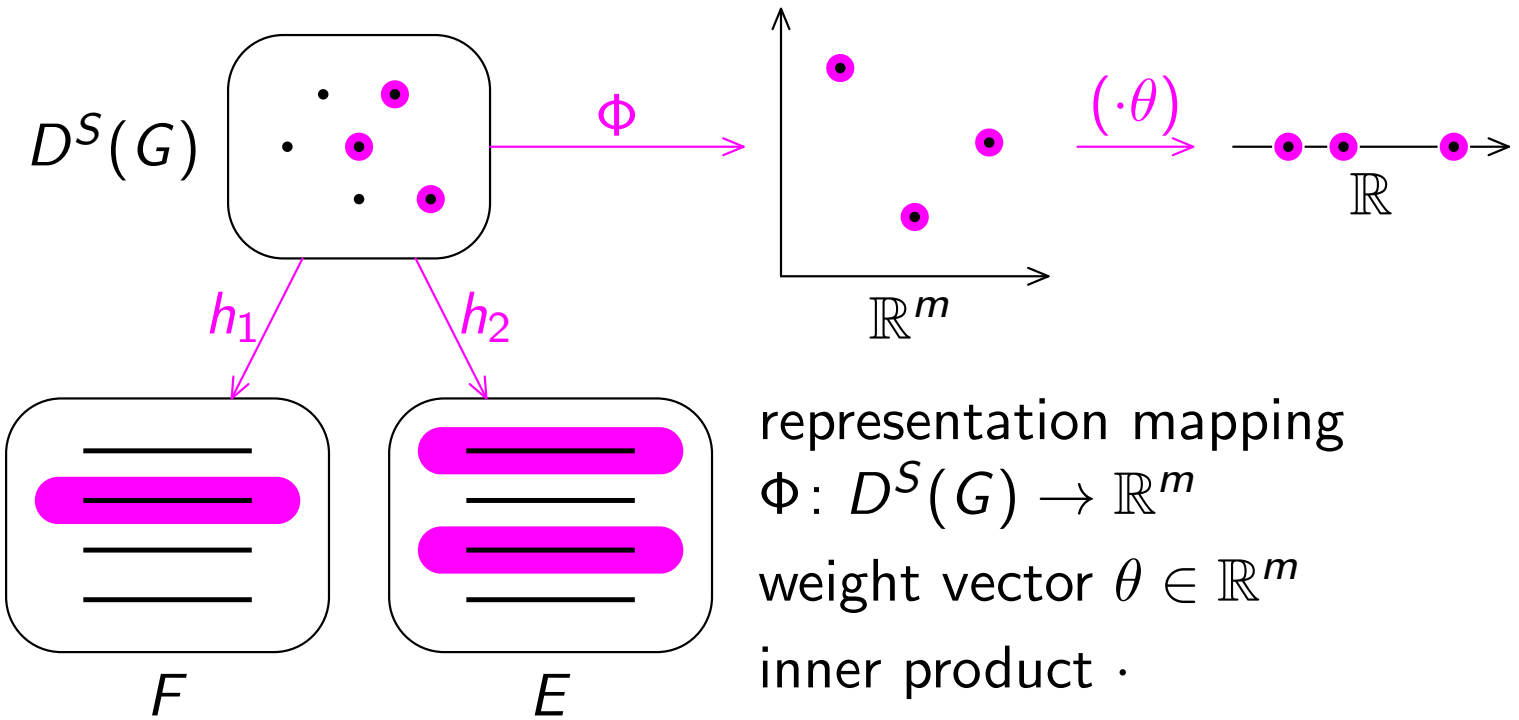


Abstract Specification

(Chiang 2007)

$\Omega = \{(G, \mu, \theta) \mid G \text{ SCFG}, \mu \text{ probability assignment}, \theta \in \mathbb{R}^3\}$

$\mathbb{D}(G, \mu, \theta)\colon f \mapsto h_2(\text{select}_{G,\mu,\theta}(\{d \mid d \in D^S(G), h_1(d) = f\}))$



representation mapping $\Phi\colon D^S(G) \to \mathbb{R}^m$

weight vector $\theta \in \mathbb{R}^m$

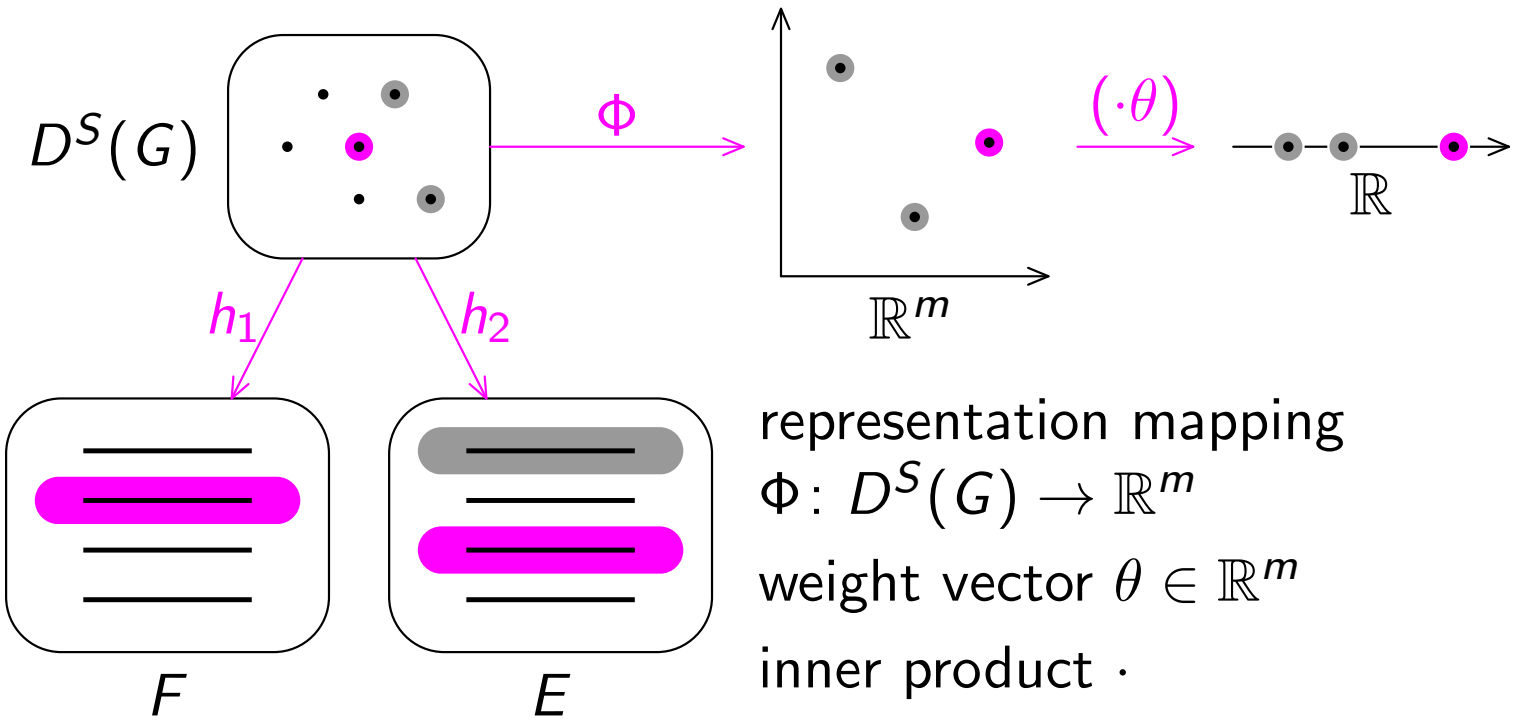
inner product $\cdot$

Abstract Specification

(Chiang 2007)

$\Omega = \{(G, \mu, \theta) \mid G \text{ SCFG}, \mu \text{ probability assignment}, \theta \in \mathbb{R}^3\}$

$\mathbb{D}(G, \mu, \theta)\colon f \mapsto h_2(\text{select}_{G,\mu,\theta}(\{d \mid d \in D^S(G), h_1(d) = f\}))$



representation mapping $\Phi\colon D^S(G) \to \mathbb{R}^m$

weight vector $\theta \in \mathbb{R}^m$

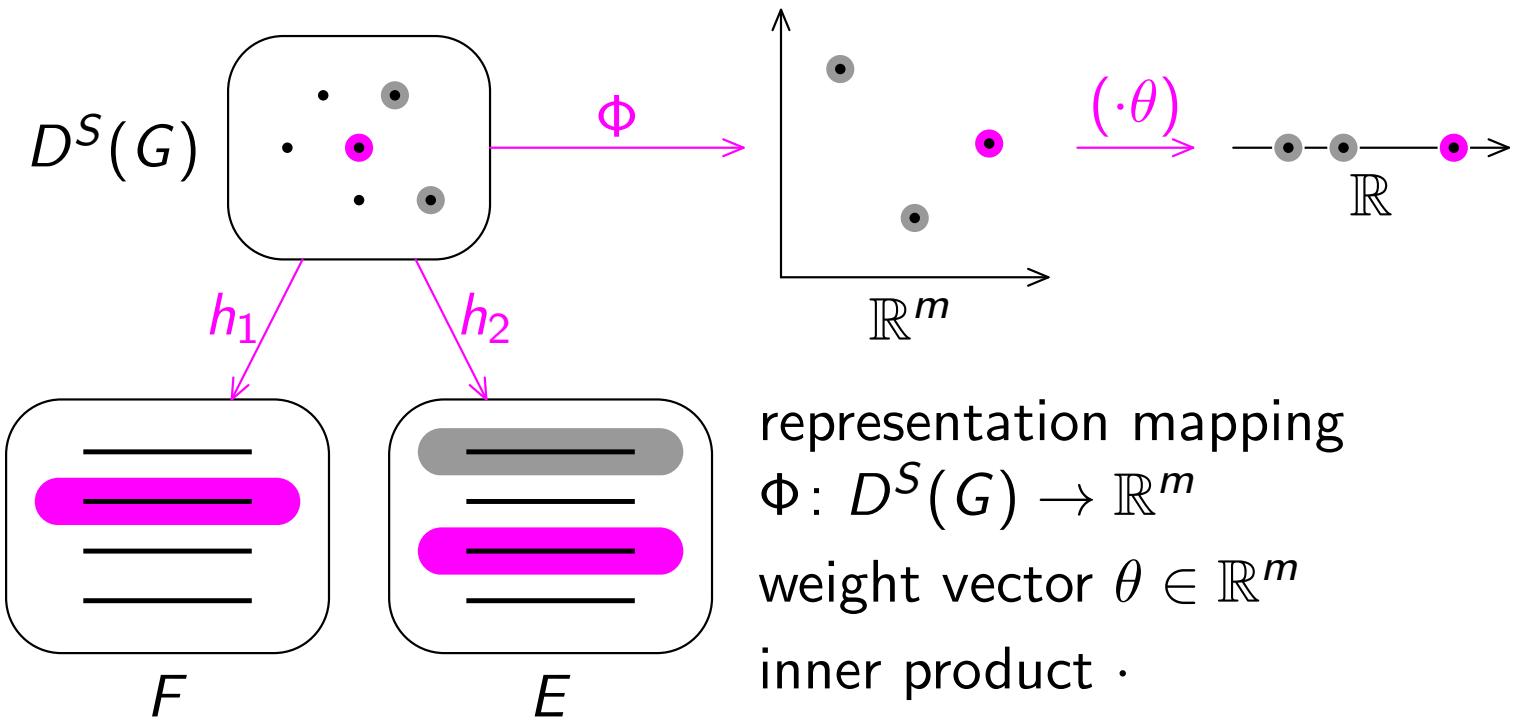
inner product $\cdot$

Abstract Specification

(Chiang 2007)

$\Omega = \{(G, \mu, \theta) \mid G \text{ SCFG}, \mu \text{ probability assignment}, \theta \in \mathbb{R}^3\}$

$\mathbb{D}(G, \mu, \theta)\colon f \mapsto h_2(\text{argmax}_{d \in D^S(G)\colon h_1(d)=f}\ \Phi_{G,\mu}(d) \cdot \theta)$



representation mapping $\Phi\colon D^S(G) \to \mathbb{R}^m$

weight vector $\theta \in \mathbb{R}^m$

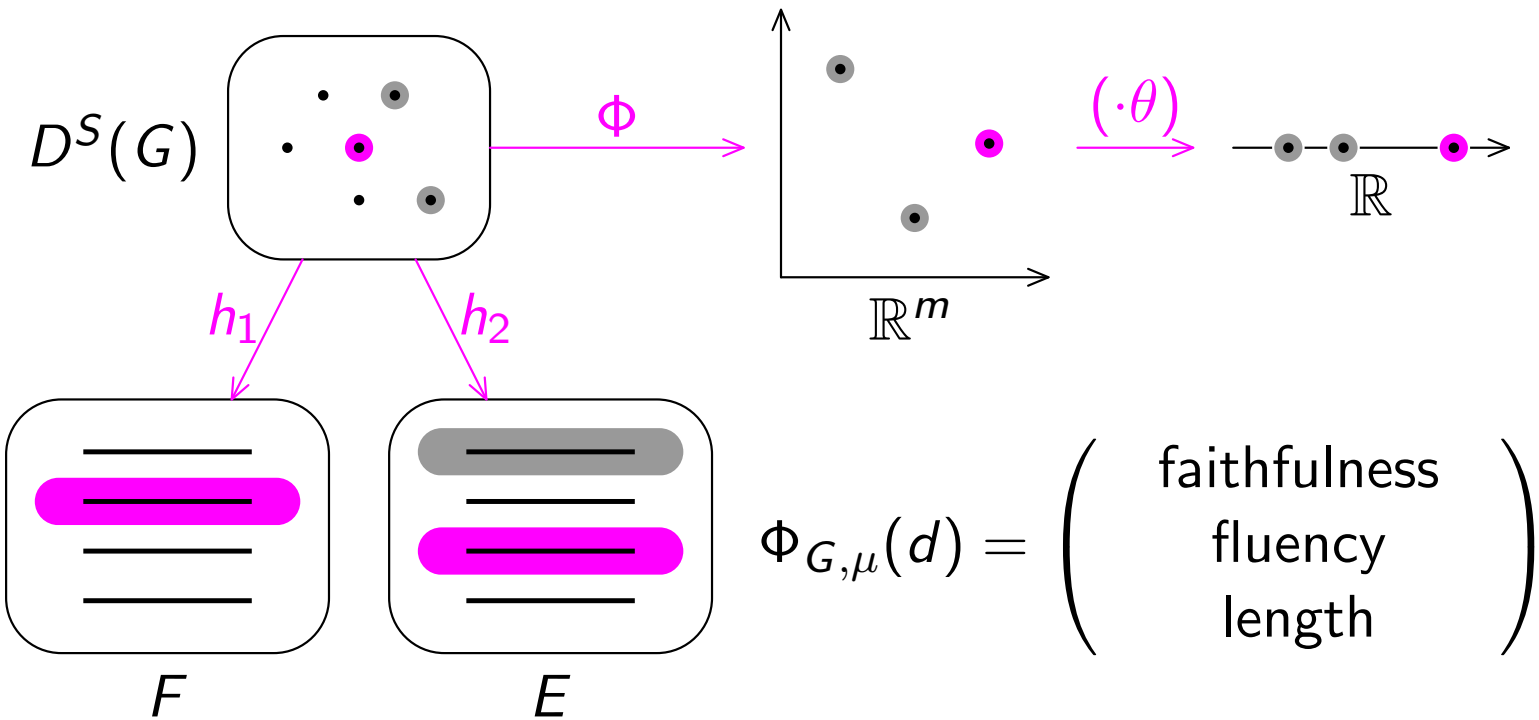
inner product $\cdot$

Abstract Specification

(Chiang 2007)

$\Omega = \{(G, \mu, \theta) \mid G \text{ SCFG}, \mu \text{ probability assignment}, \theta \in \mathbb{R}^3\}$

$\mathbb{D}(G, \mu, \theta)\colon f \mapsto h_2(\text{argmax}_{d \in D^S(G)\colon h_1(d)=f}\ \Phi_{G,\mu}(d) \cdot \theta)$



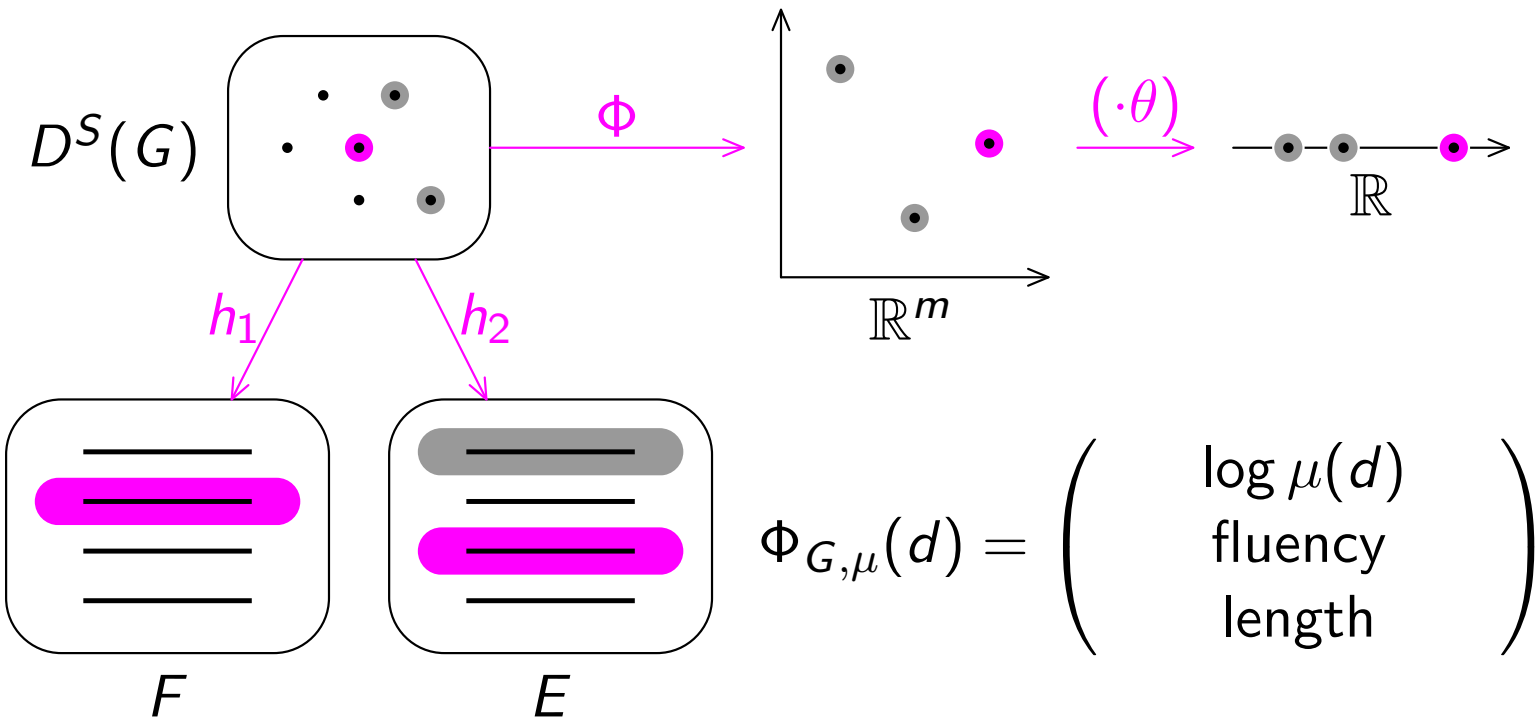
$$\Phi_{G,\mu}(d) = \begin{pmatrix} \text{faithfulness} \\ \text{fluency} \\ \text{length} \end{pmatrix}$$

Abstract Specification

(Chiang 2007)

$$\Omega = \{(G, \mu, \theta) \mid G \text{ SCFG}, \mu \text{ probability assignment}, \theta \in \mathbb{R}^3\}$$

$$\mathbb{D}(G, \mu, \theta)\colon f \mapsto h_2(\operatorname{argmax}_{d \in D^S(G)\colon h_1(d)=f} \Phi_{G,\mu}(d) \cdot \theta)$$



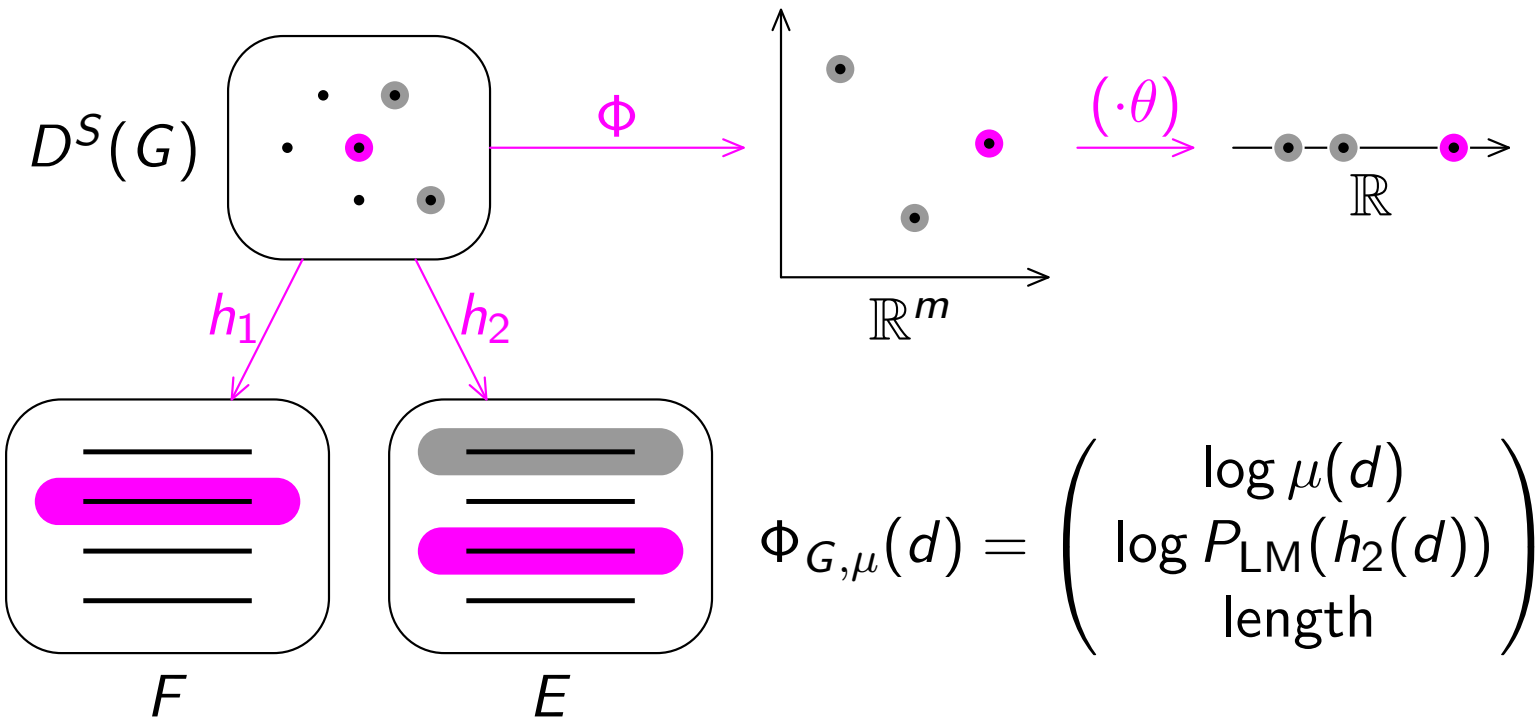
$$\Phi_{G,\mu}(d) = \begin{pmatrix} \log \mu(d) \\ \text{fluency} \\ \text{length} \end{pmatrix}$$

Abstract Specification

(Chiang 2007)

$$\Omega = \{(G, \mu, \theta) \mid G \text{ SCFG}, \mu \text{ probability assignment}, \theta \in \mathbb{R}^3\}$$

$$\mathbb{D}(G, \mu, \theta)\colon f \mapsto h_2(\operatorname{argmax}_{d \in D^S(G)\colon h_1(d)=f} \Phi_{G,\mu}(d) \cdot \theta)$$



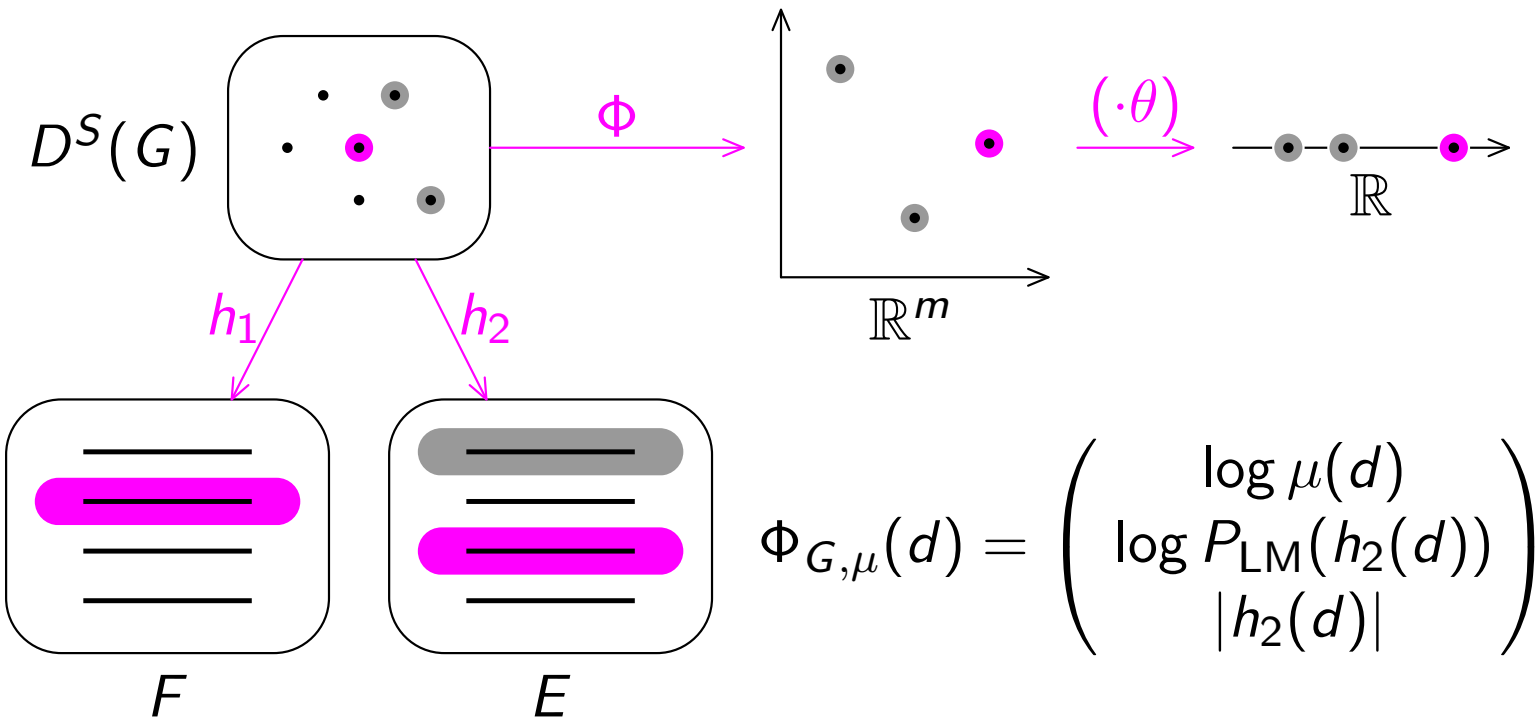
$$\Phi_{G,\mu}(d) = \begin{pmatrix} \log \mu(d) \\ \log P_{\mathrm{LM}}(h_2(d)) \\ \text{length} \end{pmatrix}$$

Abstract Specification

(Chiang 2007)

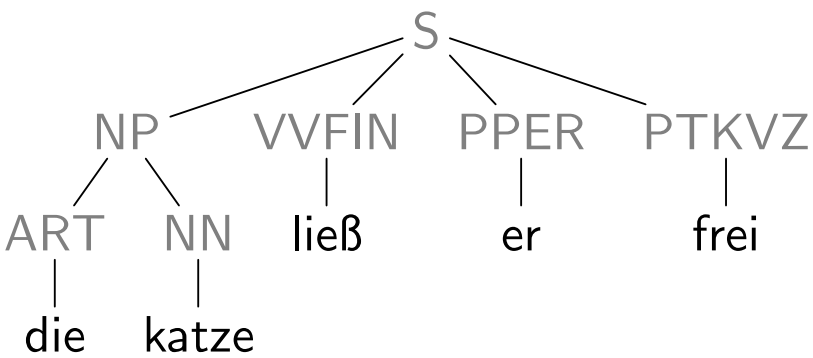
$$\Omega = \{(G, \mu, \theta) \mid G \text{ SCFG}, \mu \text{ probability assignment}, \theta \in \mathbb{R}^3\}$$

$$\mathbb{D}(G, \mu, \theta)\colon f \mapsto h_2(\operatorname{argmax}_{d \in D^S(G)\colon h_1(d)=f} \Phi_{G,\mu}(d) \cdot \theta)$$



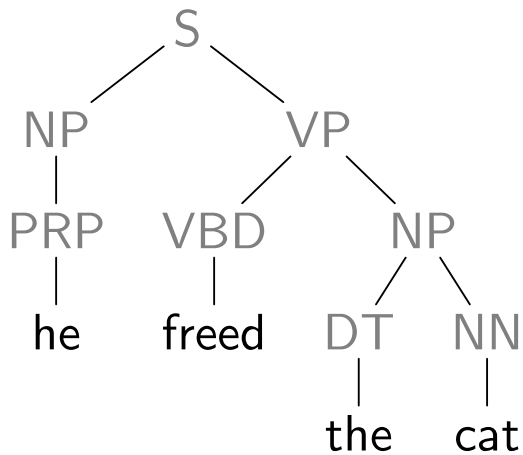
$$\Phi_{G,\mu}(d) = \begin{pmatrix} \log \mu(d) \\ \log P_{\mathrm{LM}}(h_2(d)) \\ |h_2(d)| \end{pmatrix}$$

Constituent Trees



yd

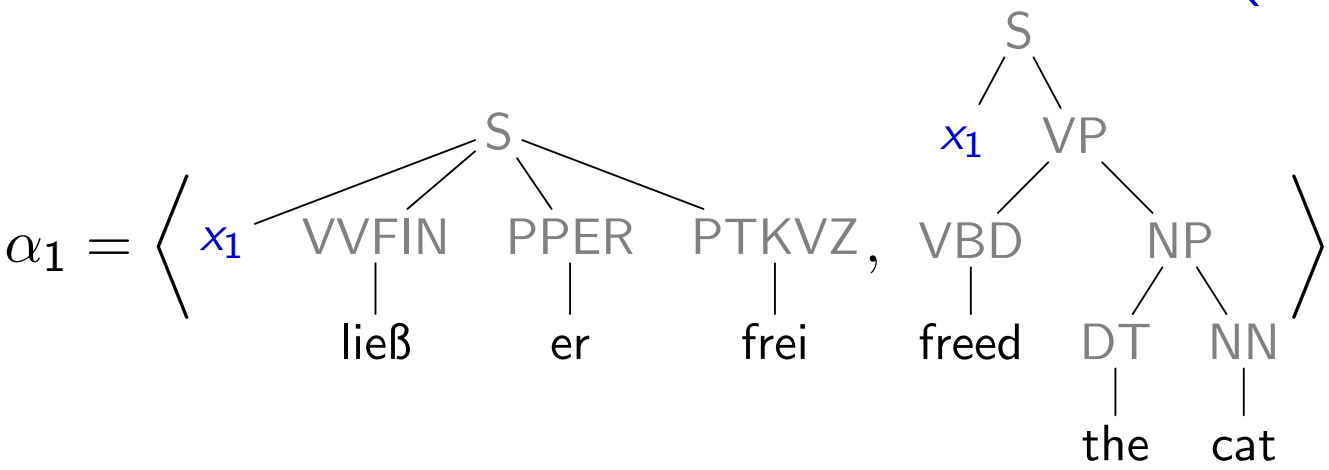
die katze ließ er frei

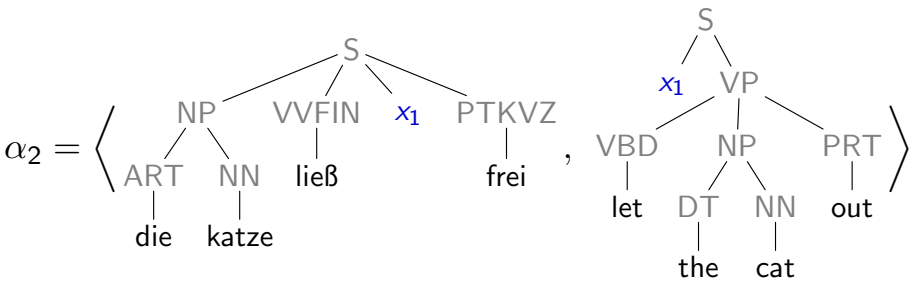


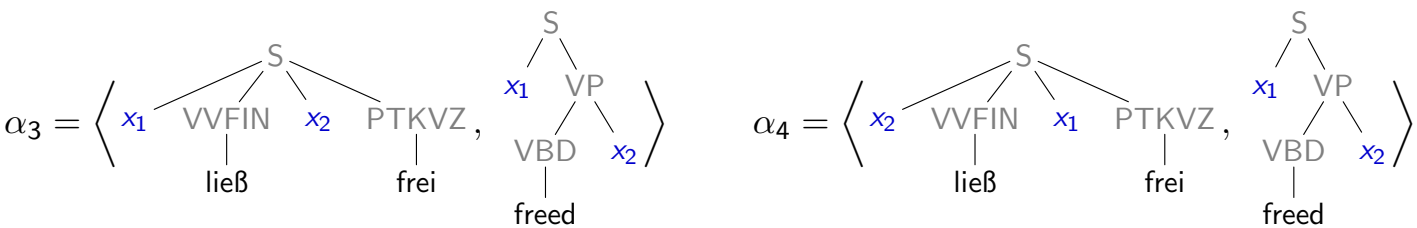
yd

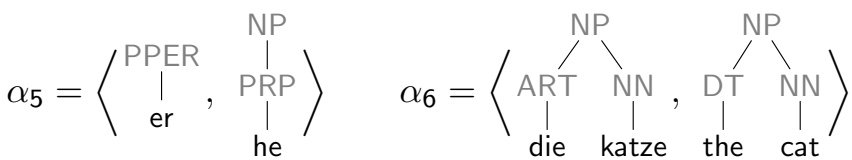
he freed the cat

Synchronous Tree-Substitution Grammar (STSG)









Decoder with Explicit Syntax (Abstract Specification)

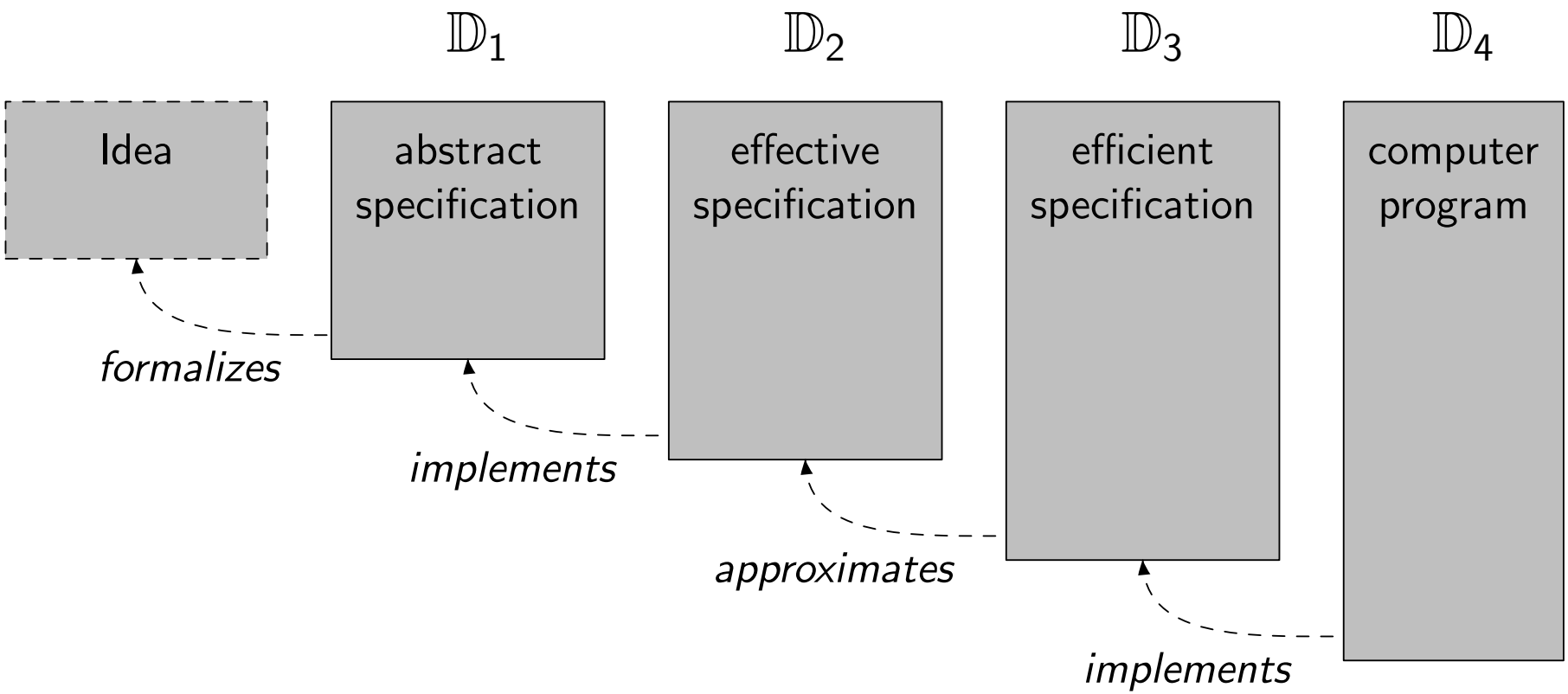
$$\Omega = \{(G, \mu, \theta) \mid G \text{ STSG}, \mu \text{ probability assignment}, \theta \in \mathbb{R}^3\}$$

$$\mathbb{D}(G, \mu, \theta)\colon f \mapsto \mathrm{yd}(h_2(\operatorname{argmax}_{d \in D^S(G)\colon\ \mathrm{yd}(h_1(d))=f} \Phi_{G,\mu}(d) \cdot \theta))$$

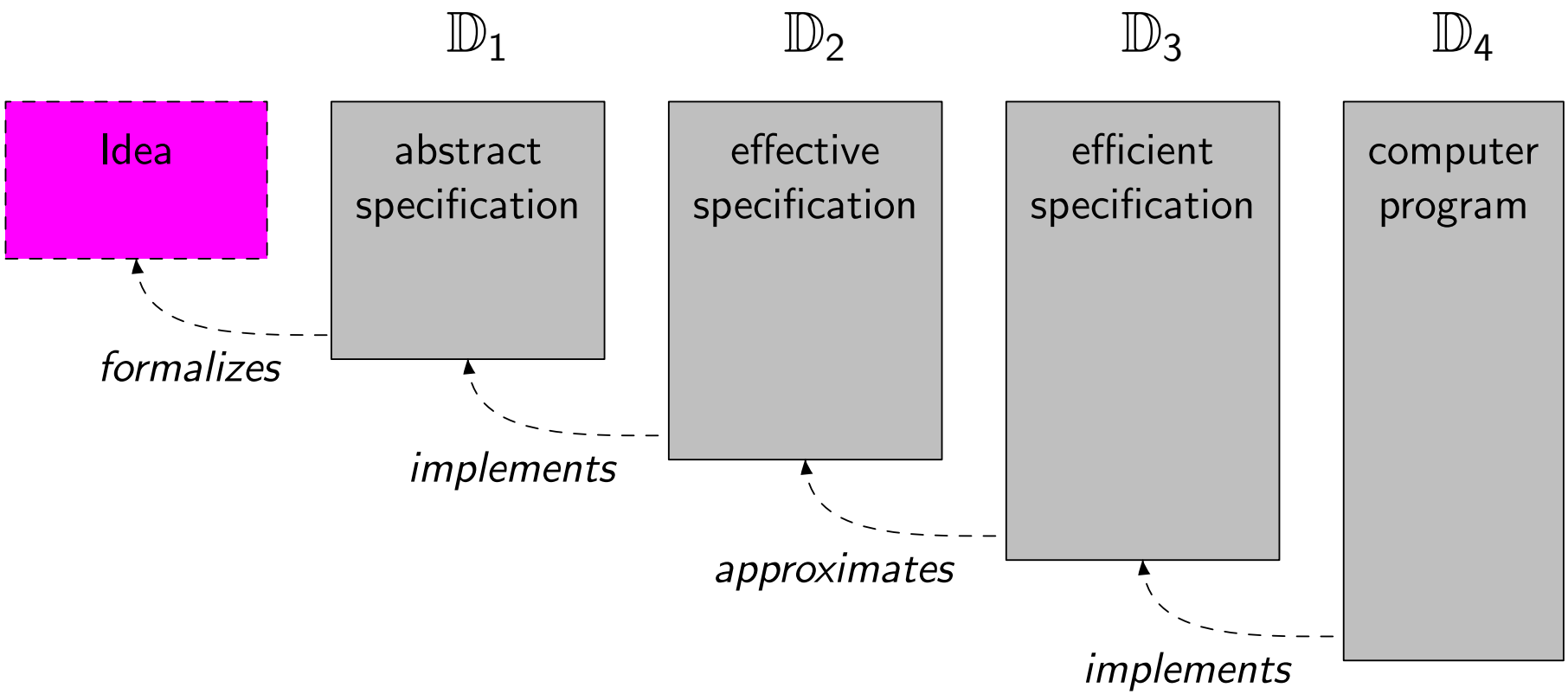
$$\Phi_{G,\mu}(d) = \begin{pmatrix} \log \mu(d) \\ \log P_{\mathrm{LM}}(\mathrm{yd}(h_2(d))) \\ \log P_{\mathrm{parse}}(h_1(d) \mid \mathrm{yd}(h_1(d))) \end{pmatrix}$$

Outline

Levels of Abstraction (Idealized)



Levels of Abstraction (Idealized)

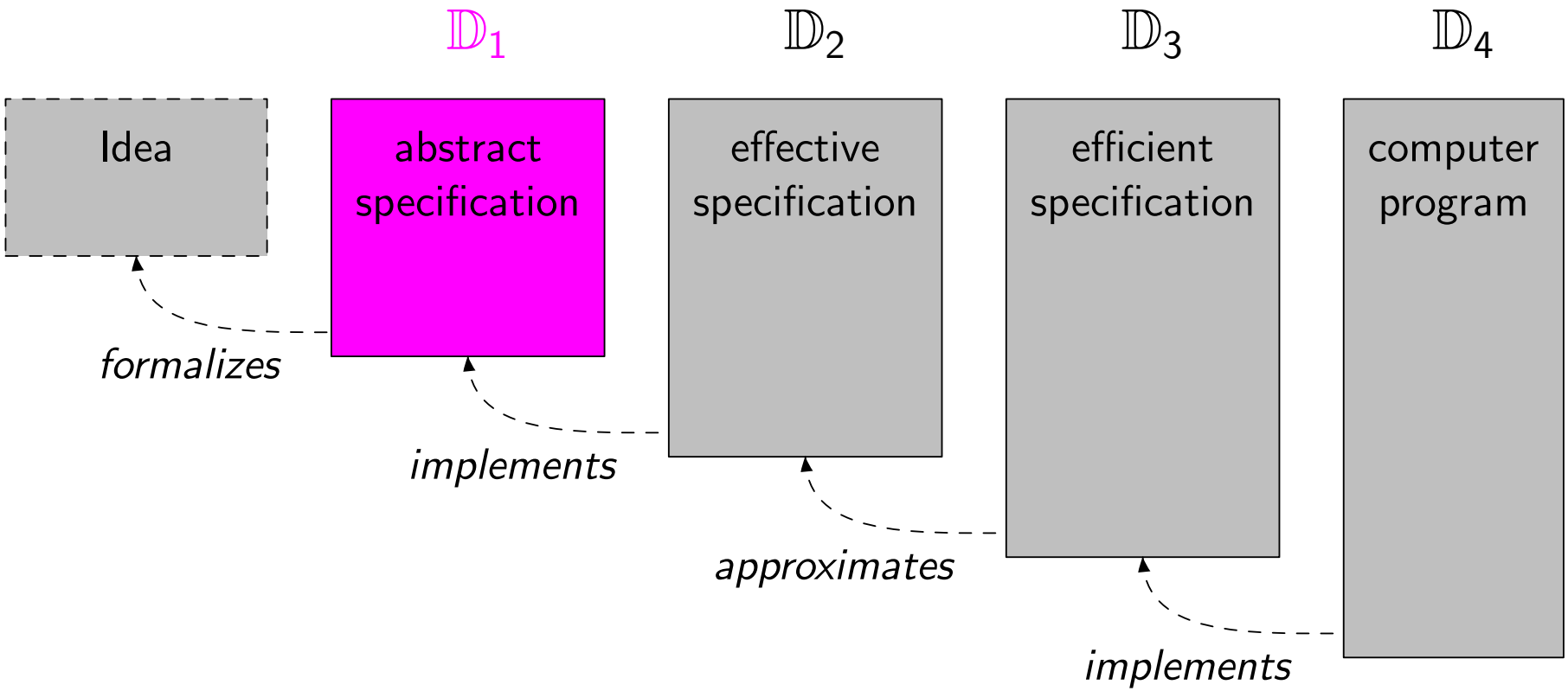


Idea

prose, examples

one page

Levels of Abstraction (Idealized)

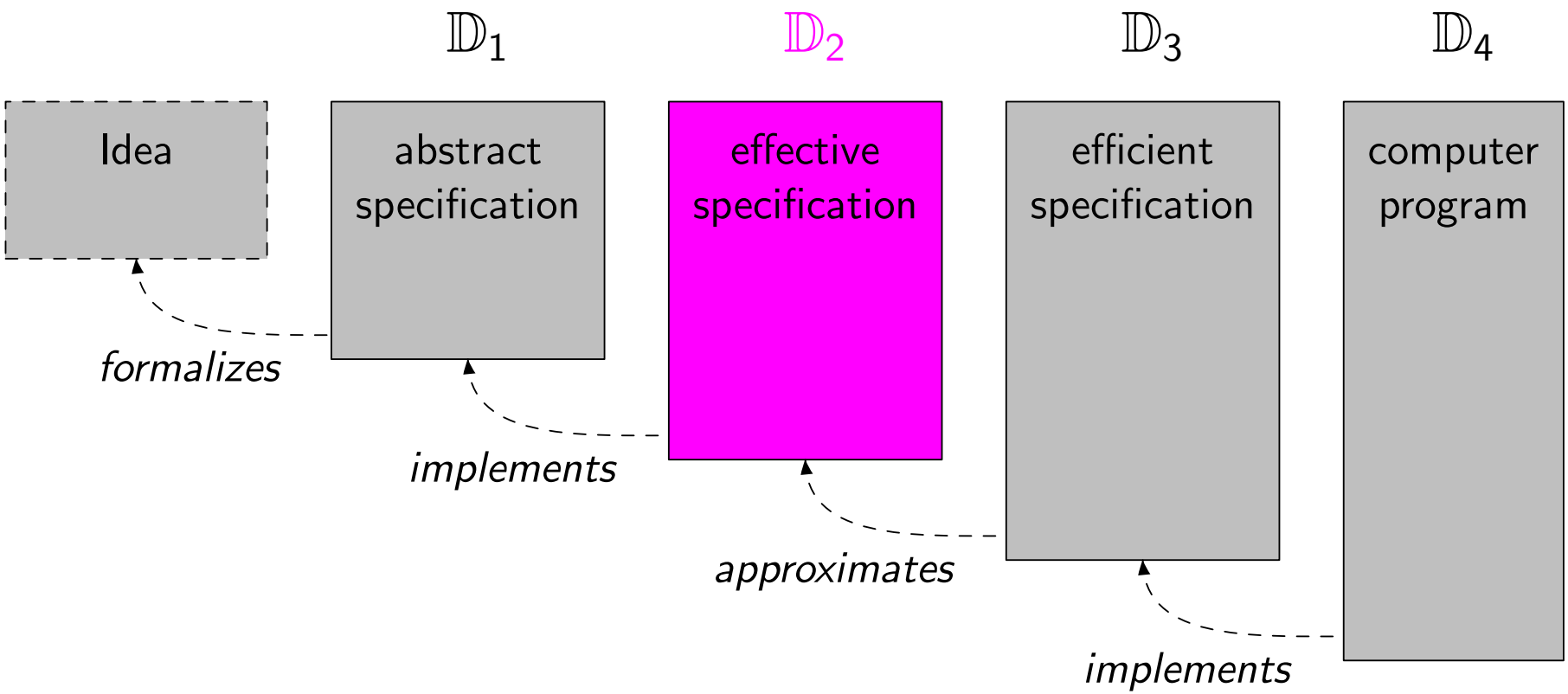


Abstract Specification

mathematical, not constructive, not unique

a few pages

Levels of Abstraction (Idealized)

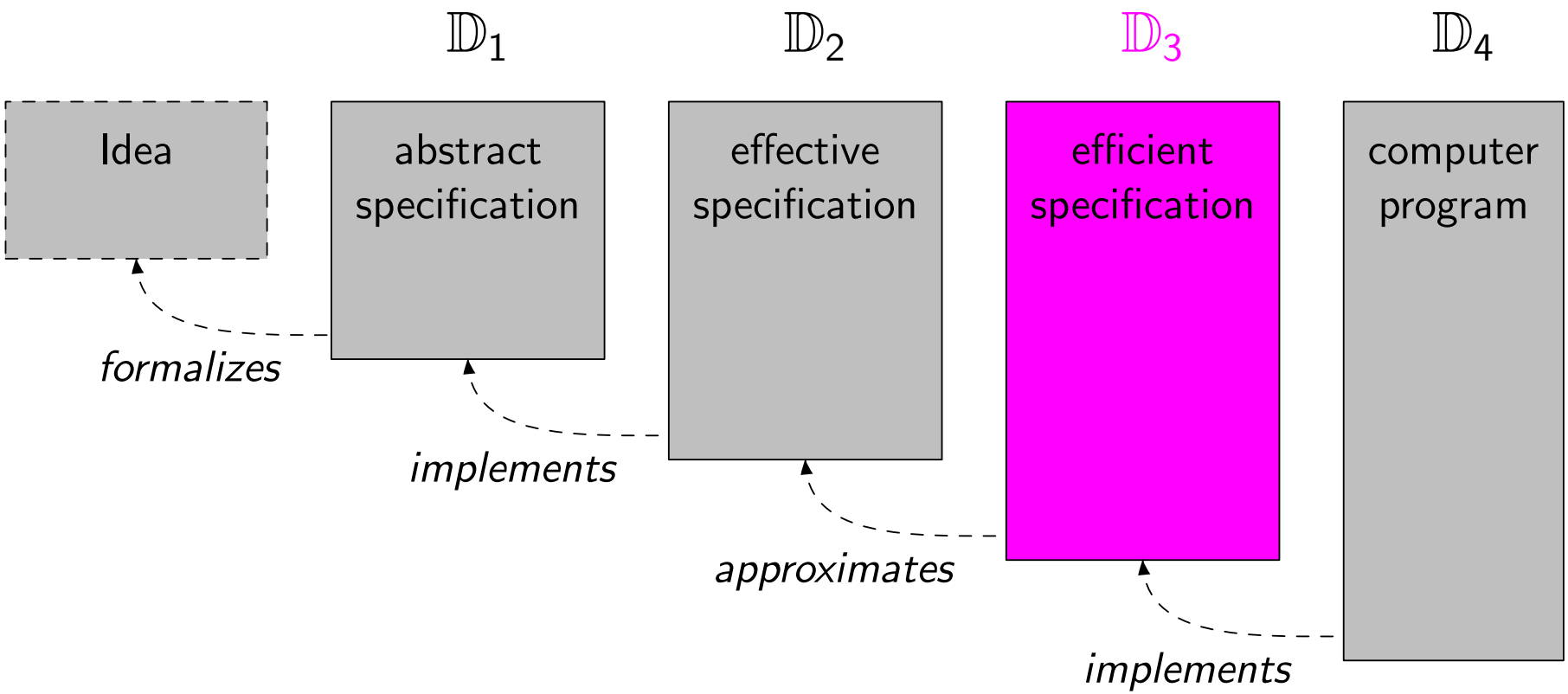


Effective Specification

mathematical, suggests an algorithm, inefficient, not unique

a slim booklet

Levels of Abstraction (Idealized)

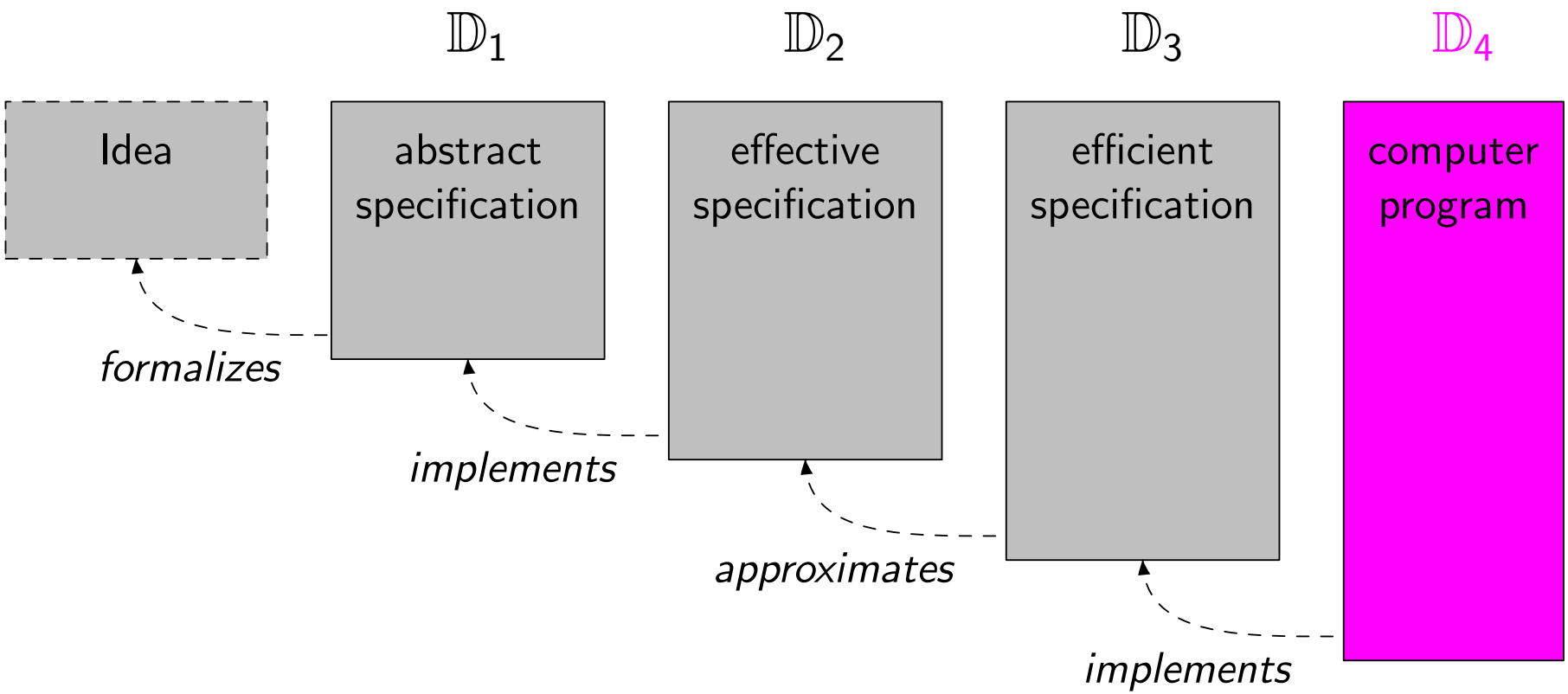


Efficient Specification

mathematical, suggests an algorithm, efficient, unique

a (very) thick book

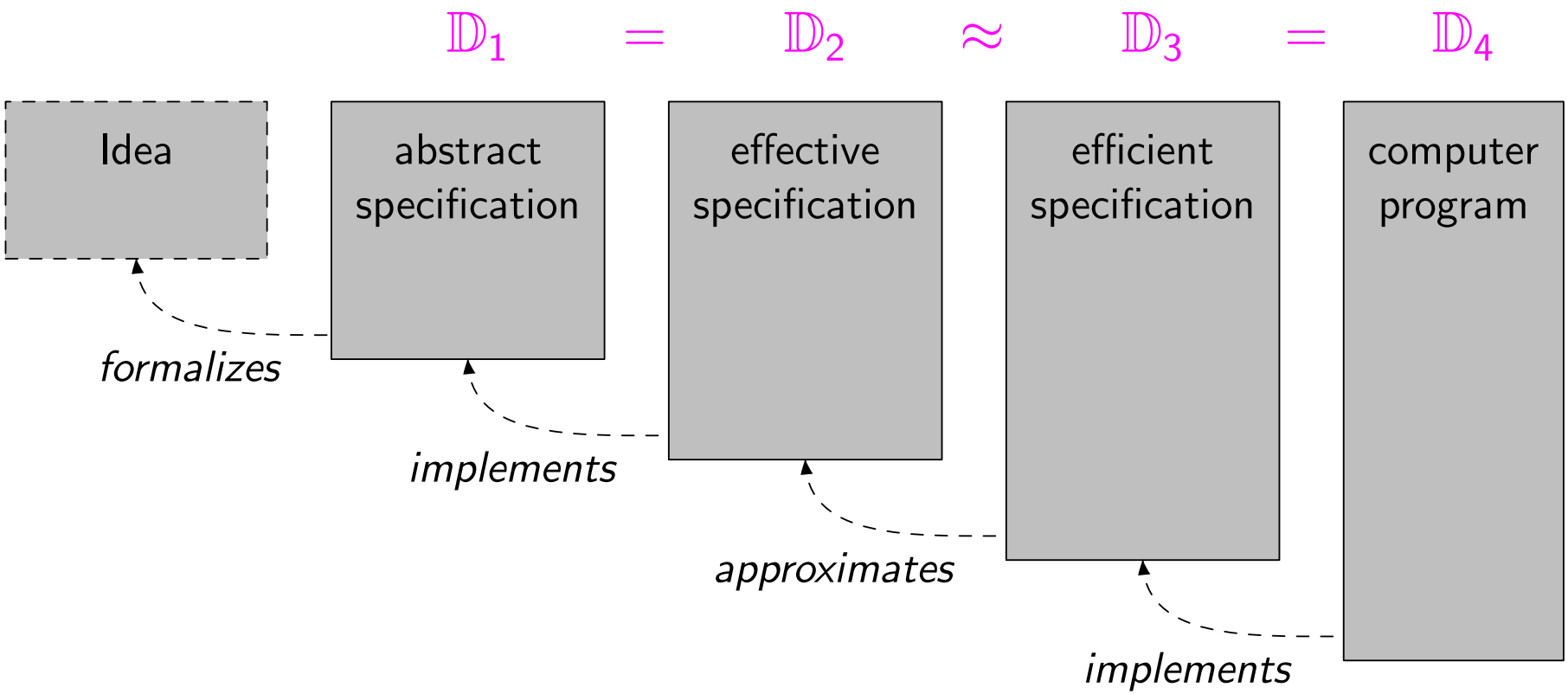
Levels of Abstraction (Idealized)



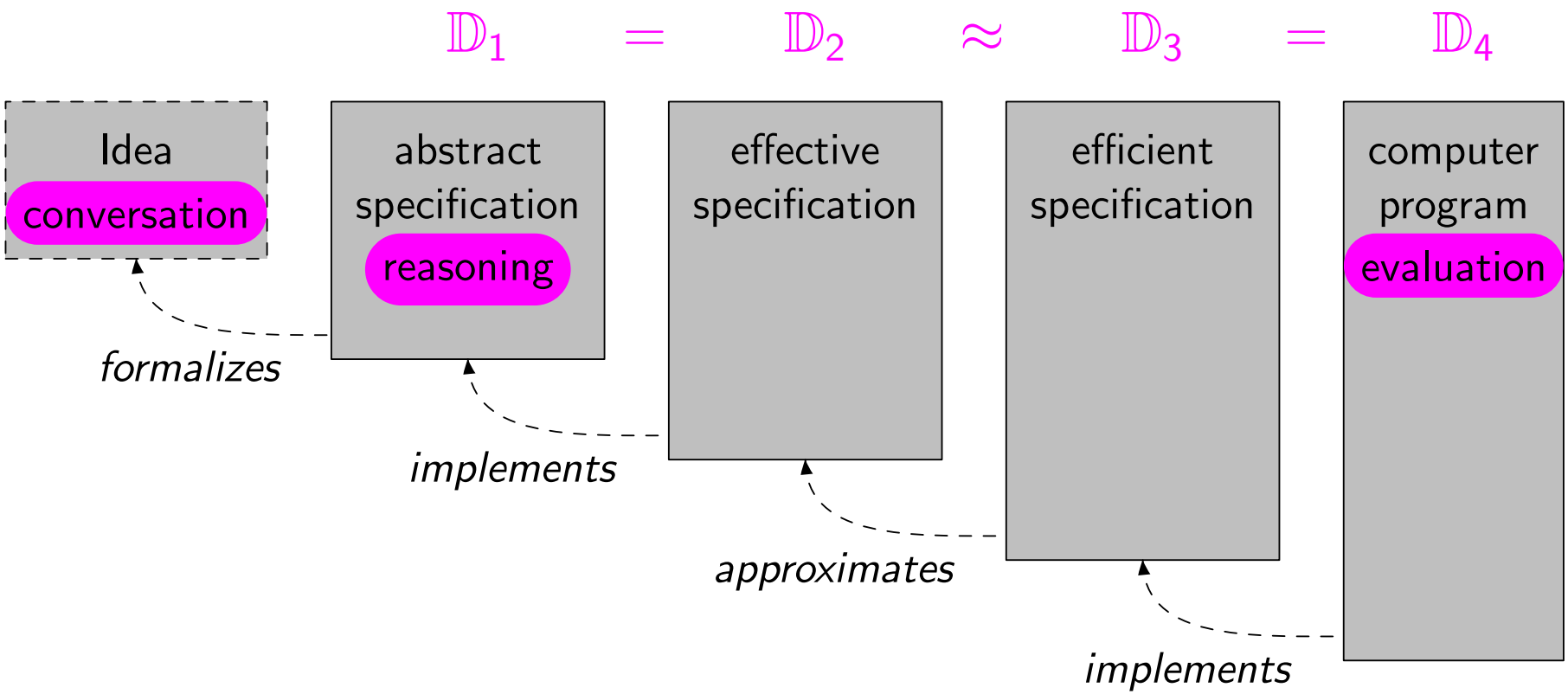
Computer Program

hundreds of thousands of lines of code (C++, Java, Python, Haskell, . . .)

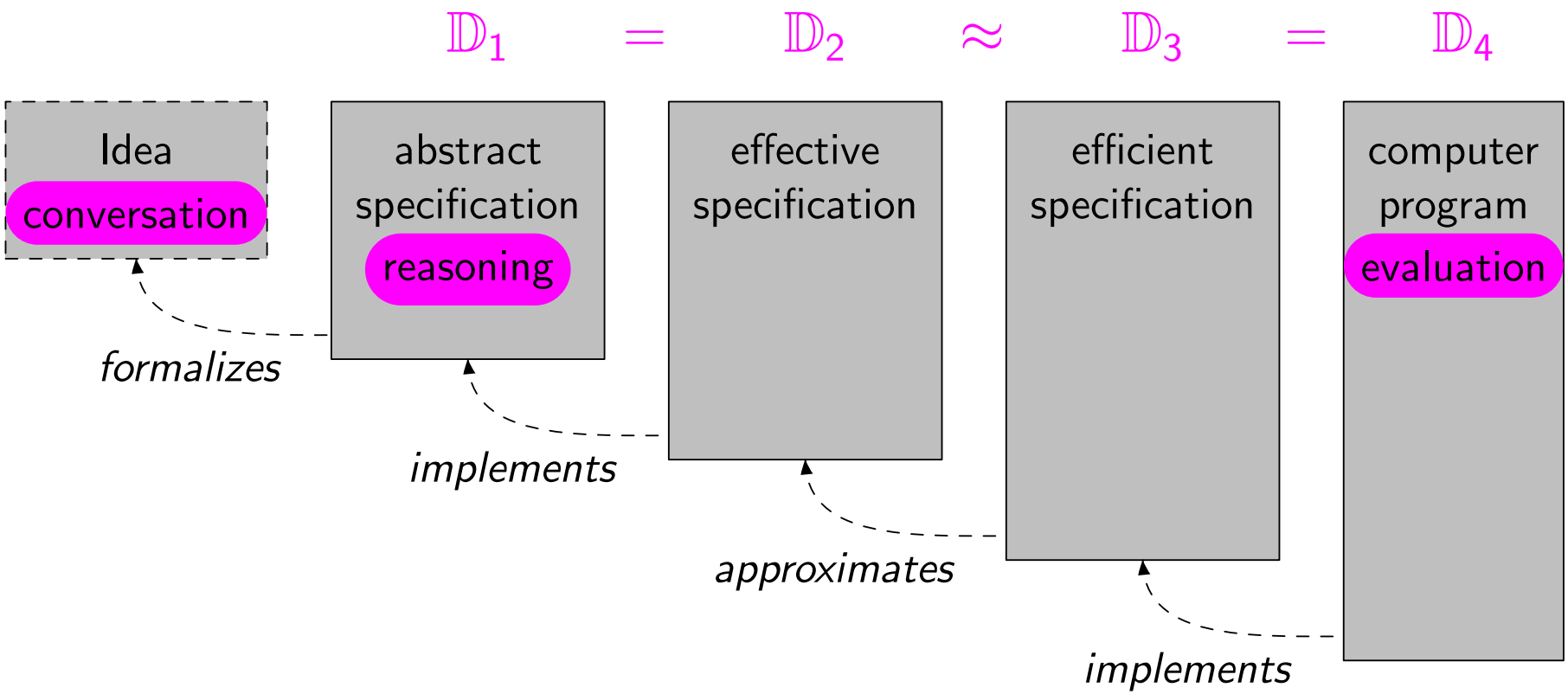
Levels of Abstraction (Idealized)



Levels of Abstraction (Idealized)



Levels of Abstraction (Idealized)

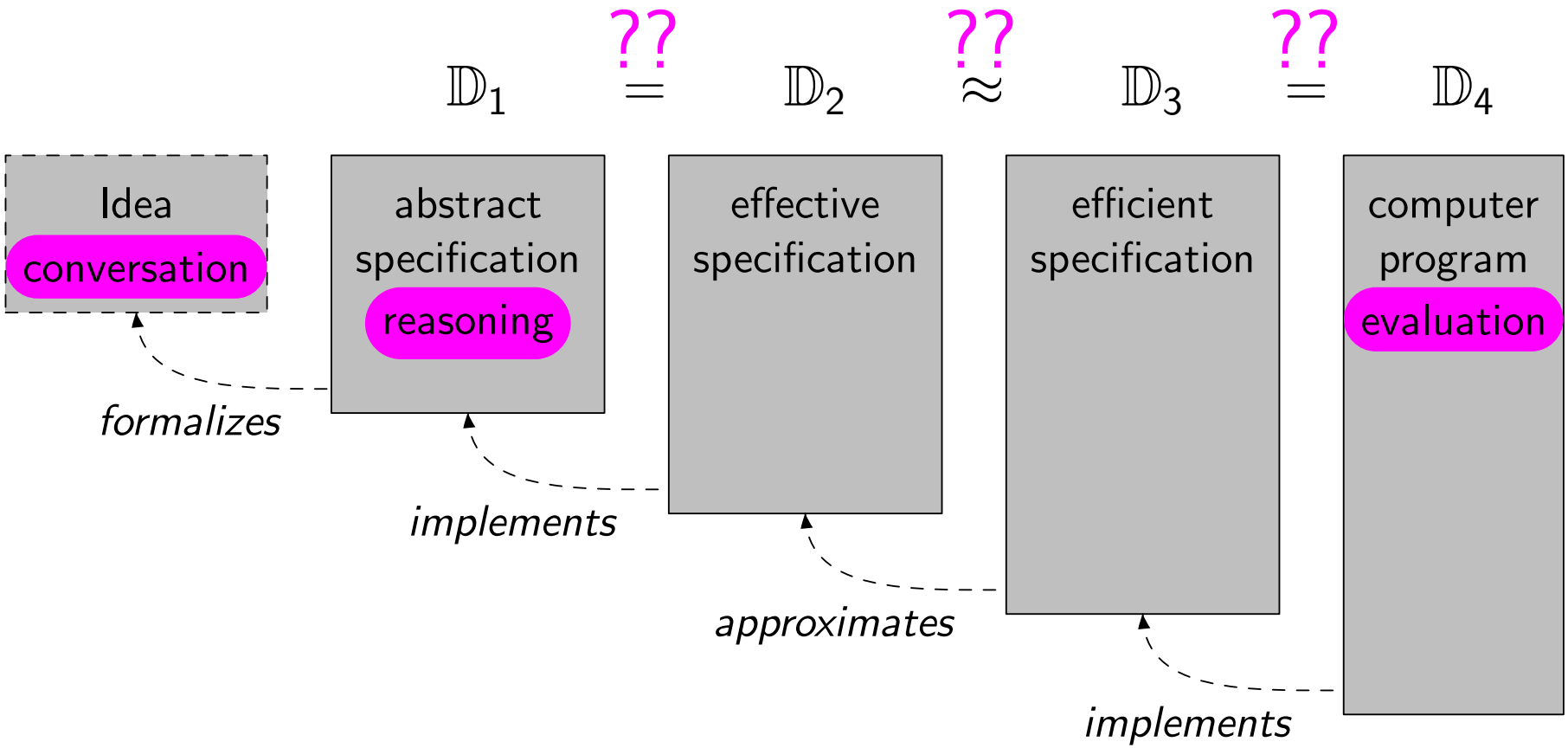


Observation

approximation impairs translation quality (but only a little)

(Chang und Collins 2011; Chiang 2007; Rush und Collins 2011)

Levels of Abstraction (Idealized)

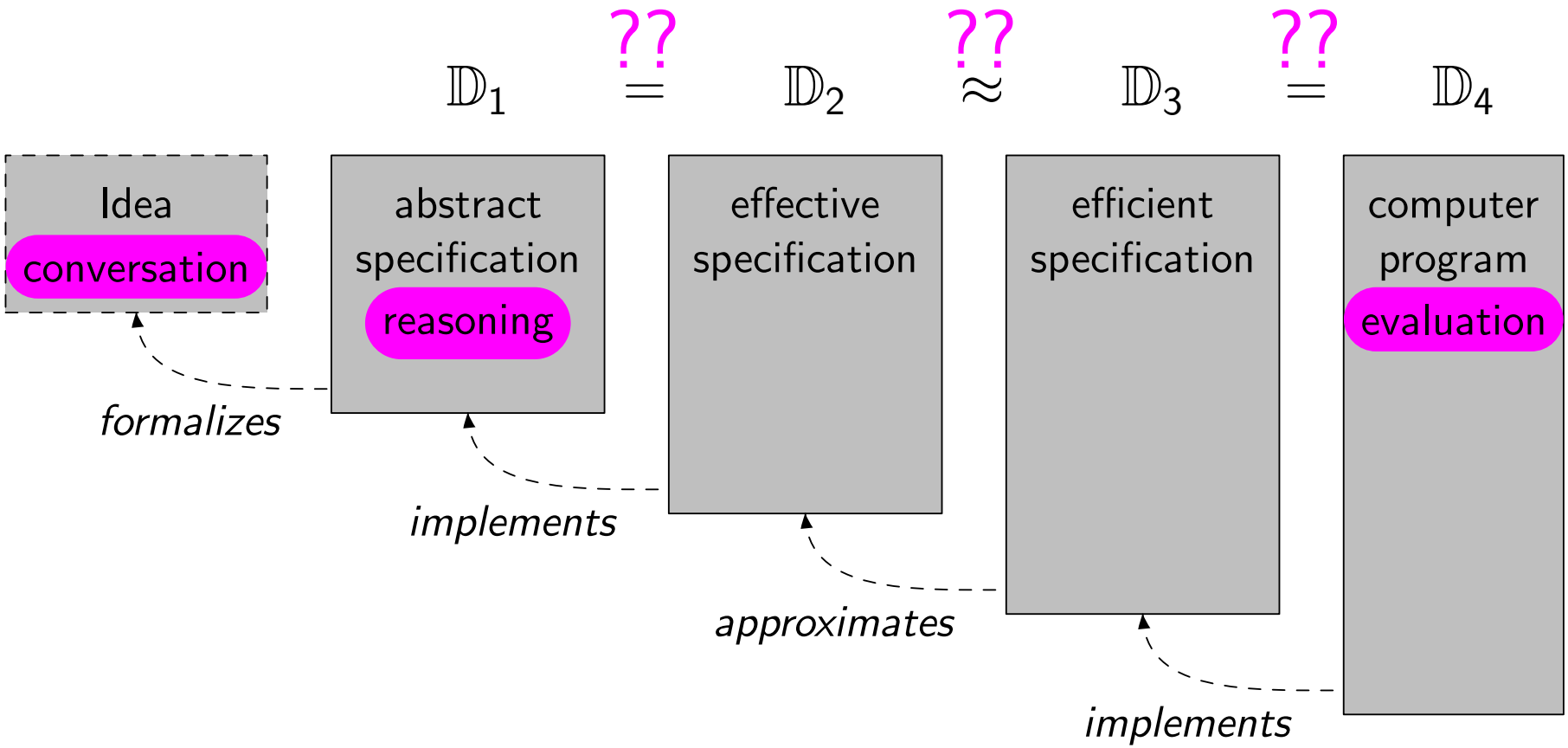


Reality

$\mathbb{D}_2$, $\mathbb{D}_3$ operational, sketchy

relationships are only being suggested

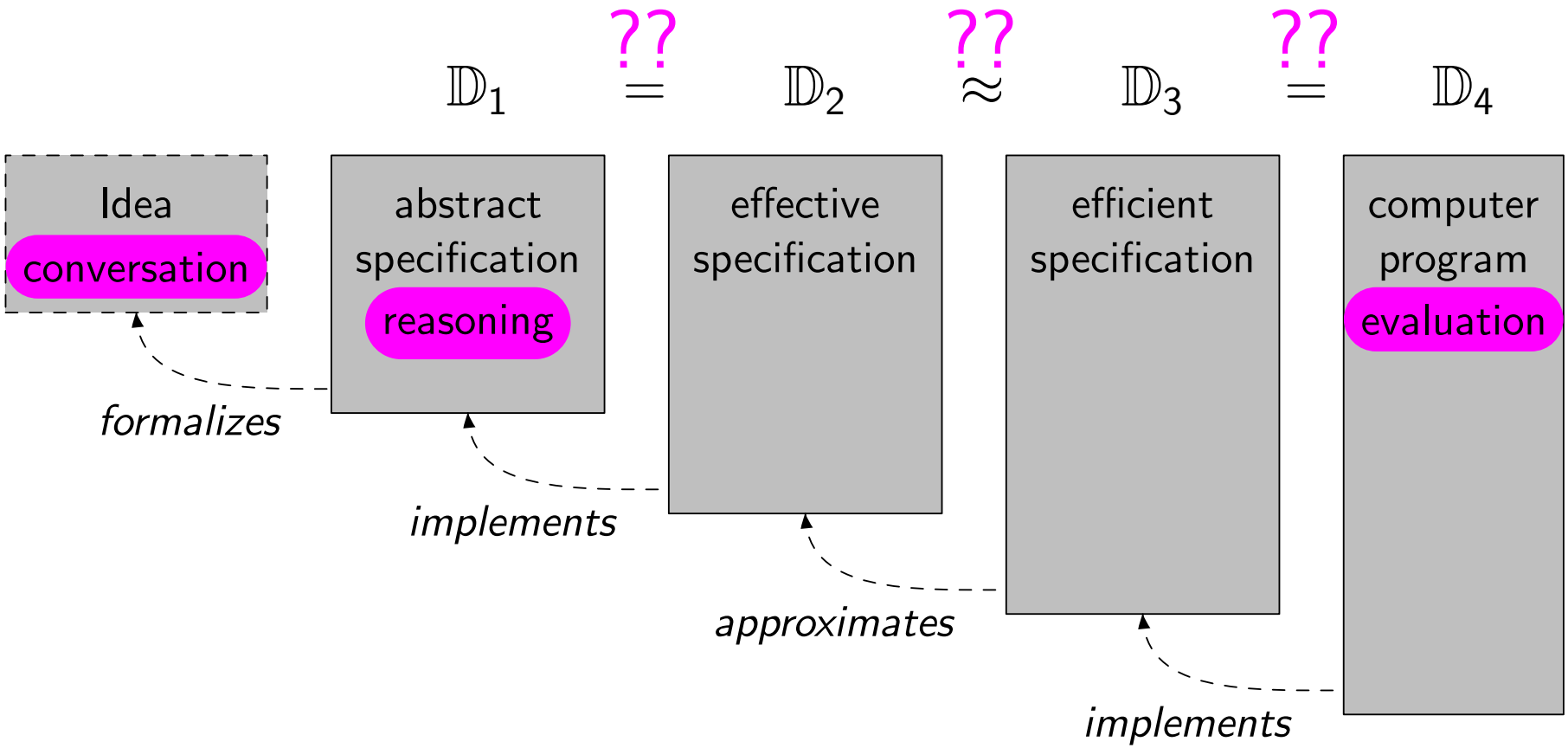
Levels of Abstraction (Idealized)



Problem

complicated, obscure, prone to errors,
hardly reusable, hard to learn!

Levels of Abstraction (Idealized)



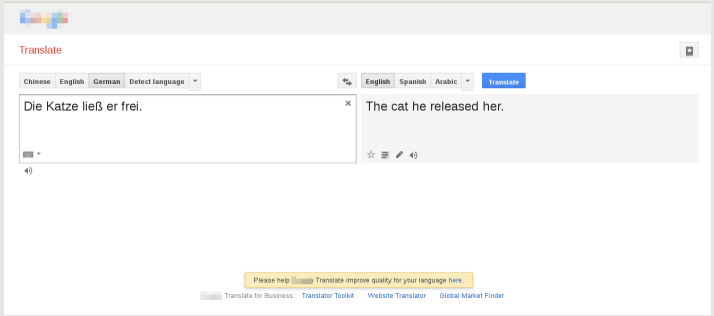
Our objective

To specify decoders better!

Not: To develop good decoders.

Overview

Statistical Machine Translation

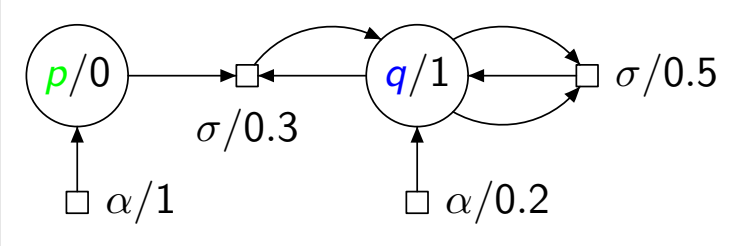


✓ 2

Decoder specification

algebraic decoder specification

Formal-Language Theory



3

closure properties

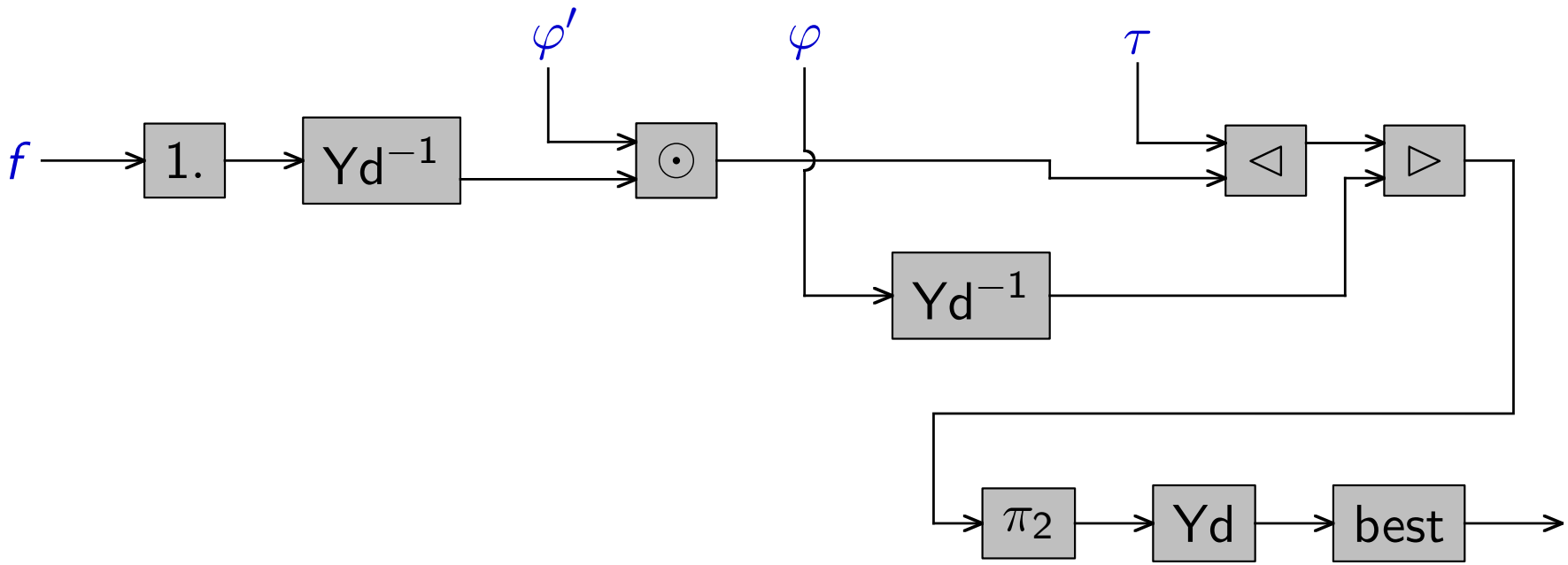
binarization

determinization

STSG-based Decoder

$\Omega = \mathcal{T}_{\mathsf{STSG}} \times \mathcal{K}_{\mathsf{Rec}} \times \mathcal{L}_{\mathsf{Rec}}$, $\mathcal{S} = (\mathbb{R}_{\geq 0}, \max, \cdot, 0, 1)$

$\mathbb{D}(\tau, \varphi, \varphi') \colon \quad f \mapsto \mathsf{best}(\mathsf{Yd}(\pi_2((\mathsf{Yd}^{-1}(1.f) \odot \varphi') \lhd \tau \rhd \mathsf{Yd}^{-1}(\varphi))))$



$\mathcal{K} = (\mathbb{R}_{\geq 0})^{\Sigma^*}$ (weighted string languages)

$\mathcal{L} = (\mathbb{R}_{\geq 0})^{T_\Sigma}$ (weighted tree languages)

$\mathcal{T} = (\mathbb{R}_{\geq 0})^{T_\Sigma \times T_\Sigma}$ (weighted tree transformations)

Example f

die katze ließ er frei

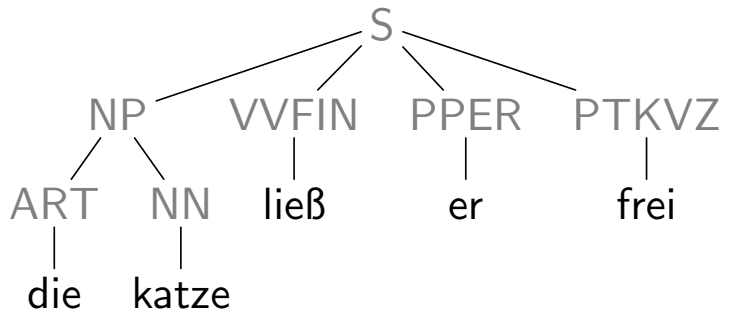
String Injection 1.: $\Sigma^* \to \mathcal{K}$

1.f:

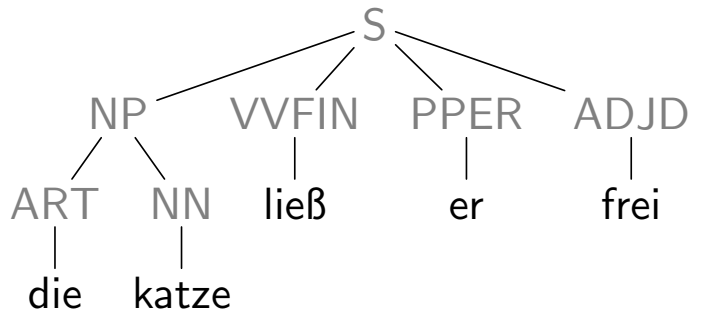
$$\begin{aligned} \ldots &\mapsto 0 \\ \text{die katze ließ er frei} &\mapsto 1 \\ \ldots &\mapsto 0 \end{aligned}$$

Language Yield Inverse $\mathrm{Yd}^{-1}\colon \mathcal{K} \to \mathcal{L}$

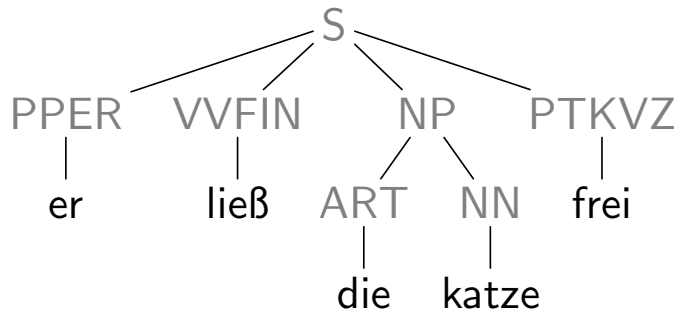
$\mathrm{Yd}^{-1}(1.f)$:



$\mapsto$ 1



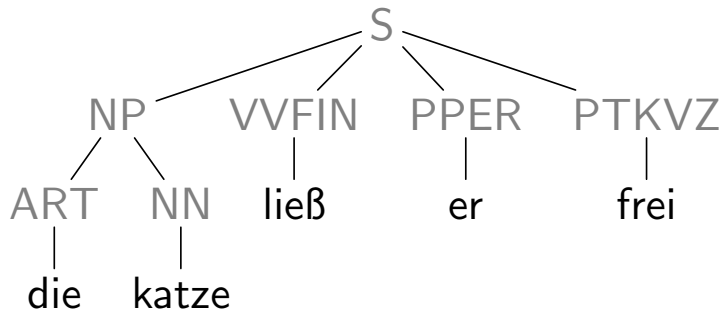
$\mapsto$ 1



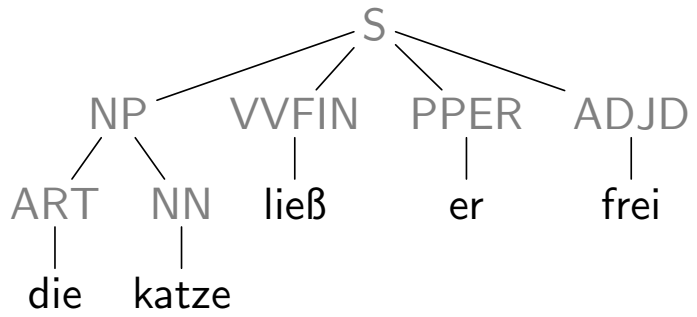
$\mapsto$ 0

$\cdots \mapsto$ 0

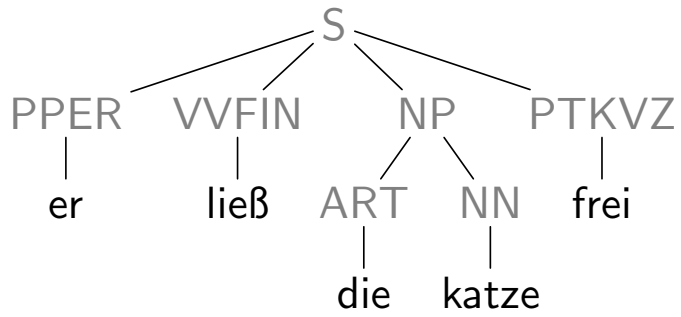
Example φ'



$\mapsto$ 0.5



$\mapsto$ 0.1

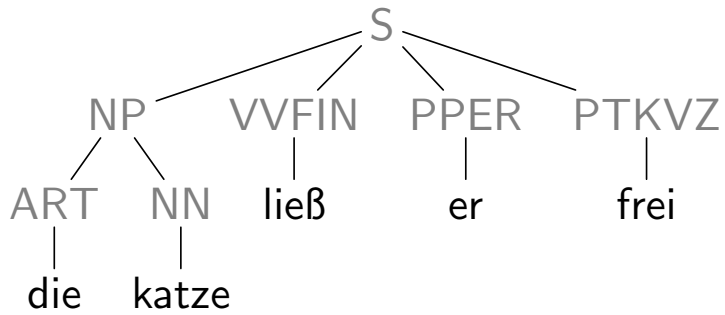


$\mapsto$ 0.2

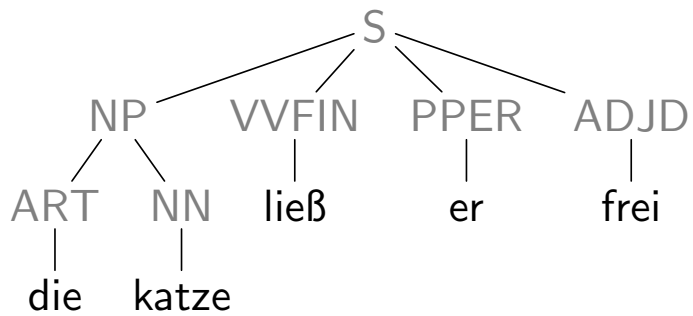
$\ldots \mapsto \ldots$

Hadamard Product $\odot\colon \mathcal{L} \times \mathcal{L} \to \mathcal{L}$

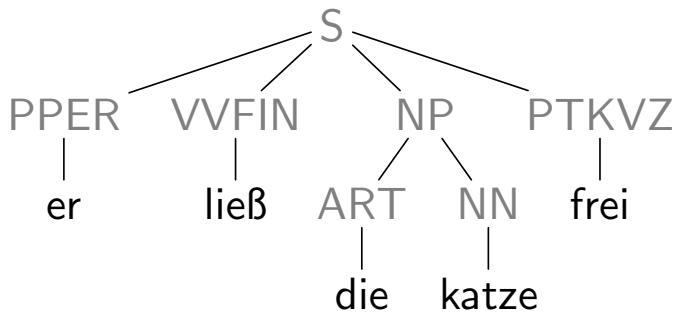
$\mathrm{Yd}^{-1}(1.f) \odot \varphi'$:



$\mapsto$ 0.5



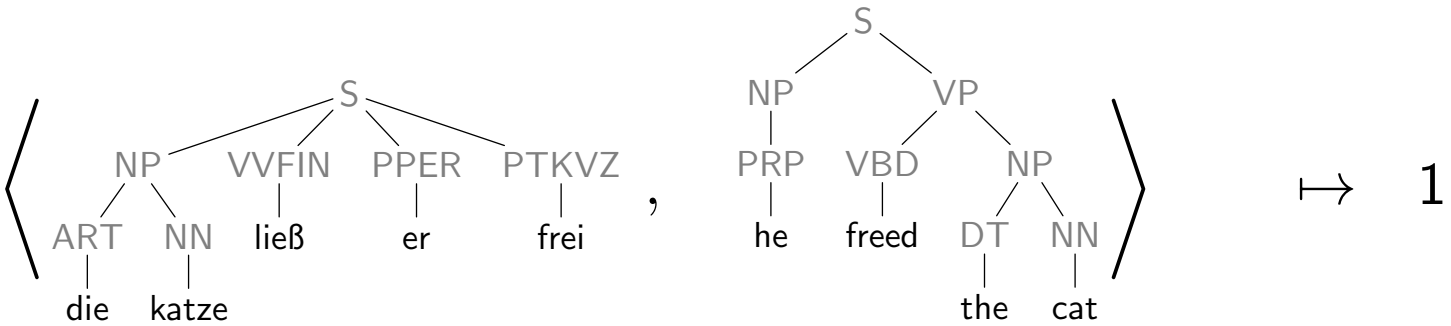
$\mapsto$ 0.1



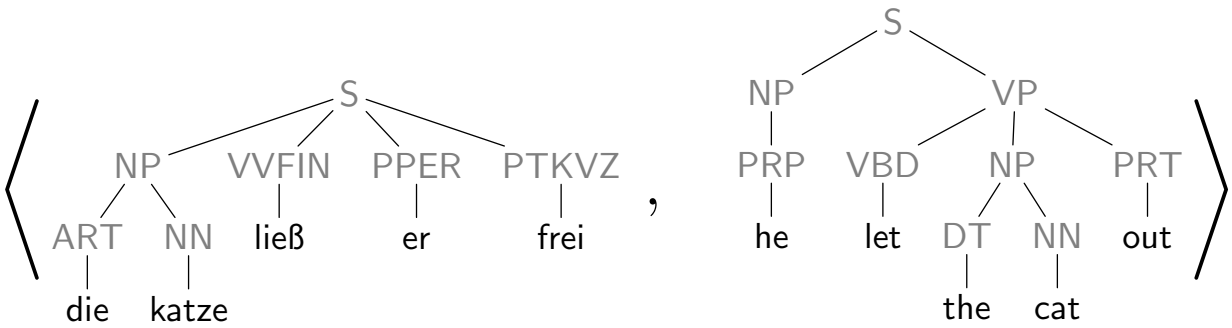
$\mapsto$ 0

$\cdots \mapsto 0$

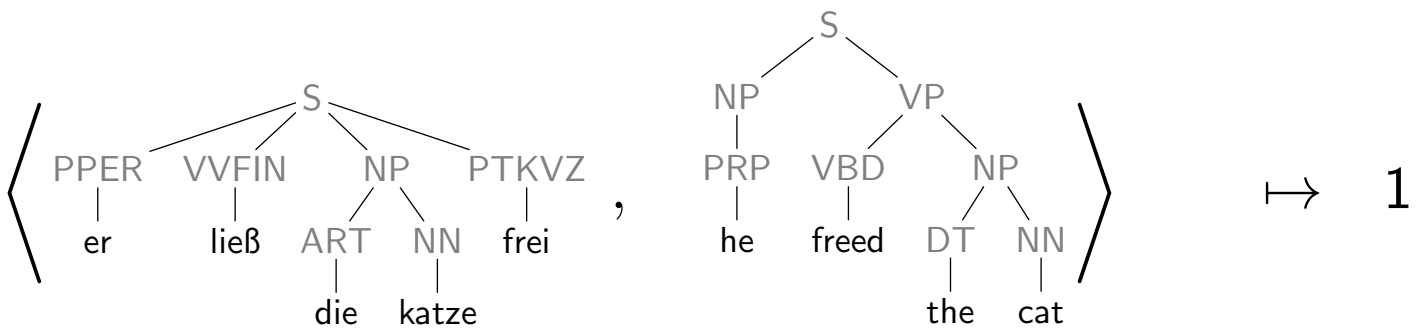
Example τ



$\mapsto$ 1



$\mapsto$ 1

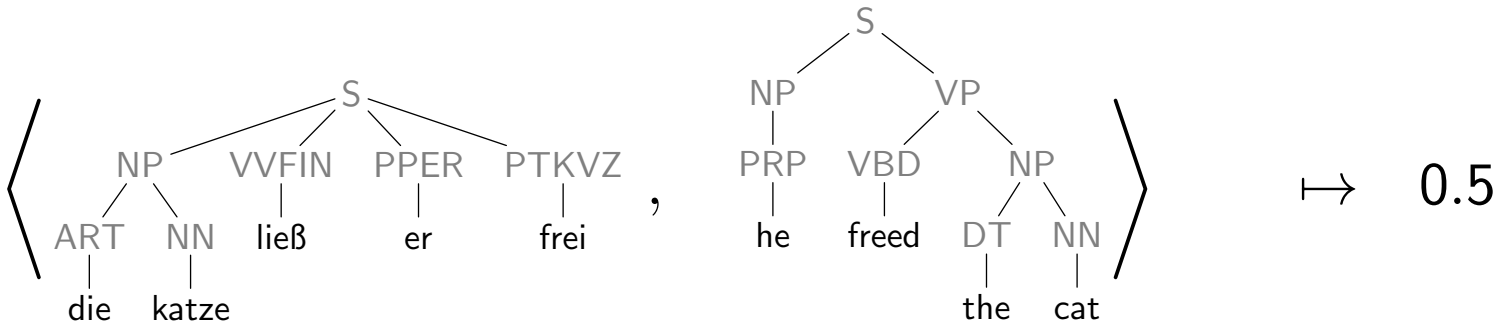


$\mapsto$ 1

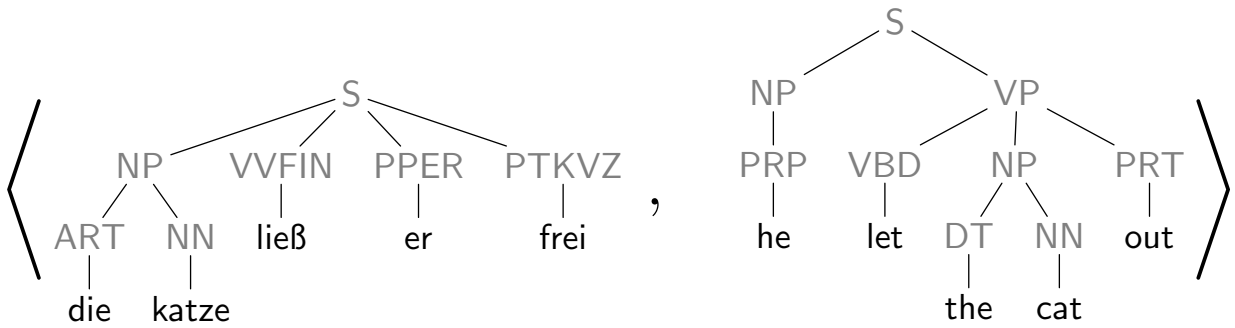
$\ldots$ $\mapsto$ 0

Input Product $\lhd\colon \mathcal{L} \times \mathcal{T} \to \mathcal{T}$

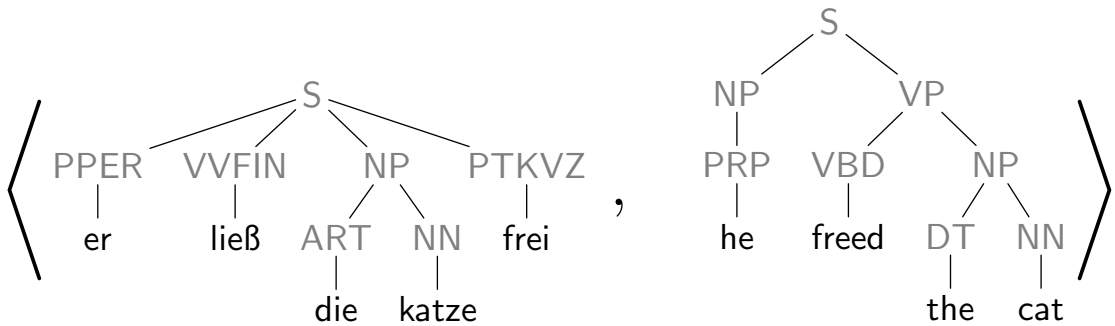
$(\mathrm{Yd}^{-1}(1.f) \odot \varphi') \lhd \tau$:



$\mapsto$ 0.5



$\mapsto$ 0.5



$\mapsto$ 0

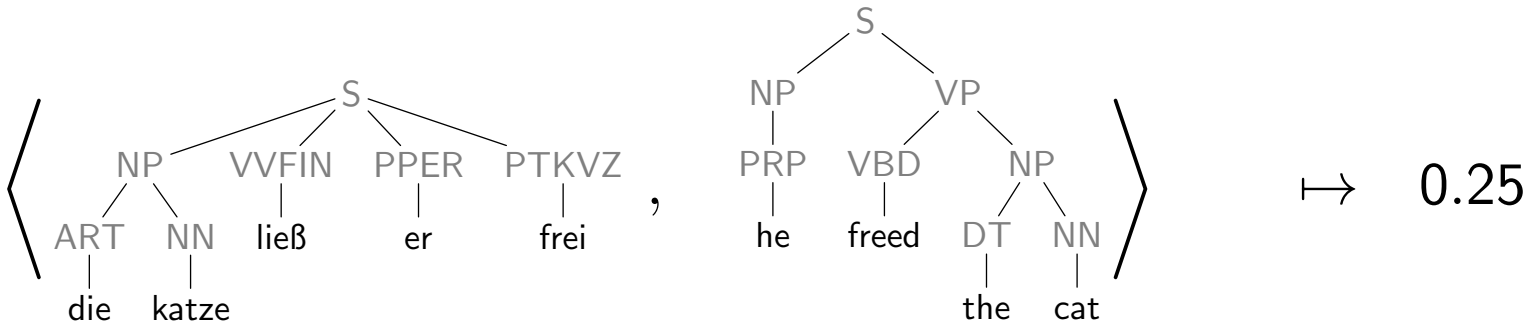
$\ldots$ $\mapsto$ 0

Example φ

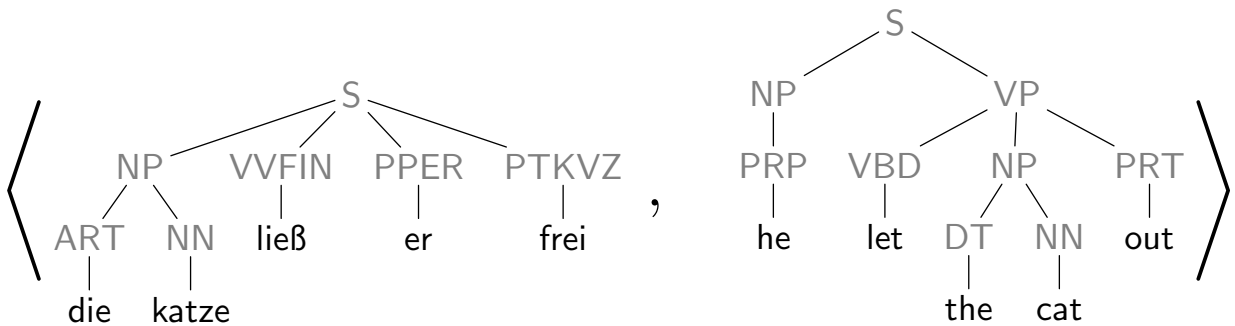
$$\begin{aligned} \text{he freed the cat} &\mapsto 0.5 \\ \text{he let the cat out} &\mapsto 0.1 \\ \ldots &\mapsto \ldots \end{aligned}$$

Output Product $\triangleright : \mathcal{T} \times \mathcal{L} \to \mathcal{T}$

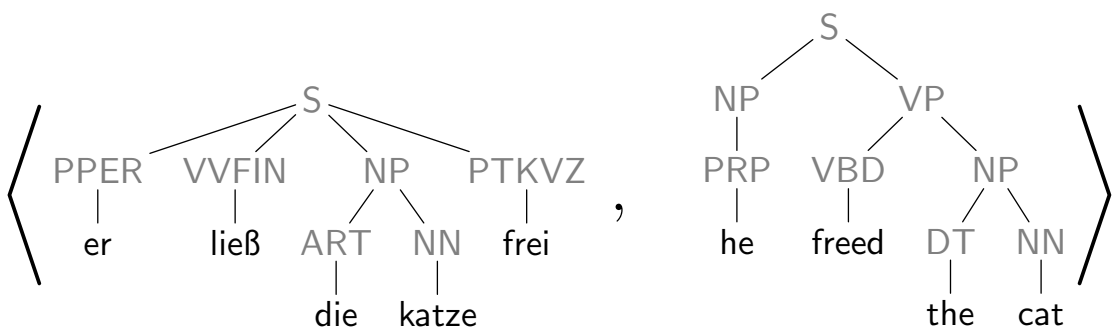
$(\mathrm{Yd}^{-1}(1.f) \odot \varphi') \lhd \tau \rhd \mathrm{Yd}^{-1}(\varphi)$:



$\mapsto$ 0.25



$\mapsto$ 0.05

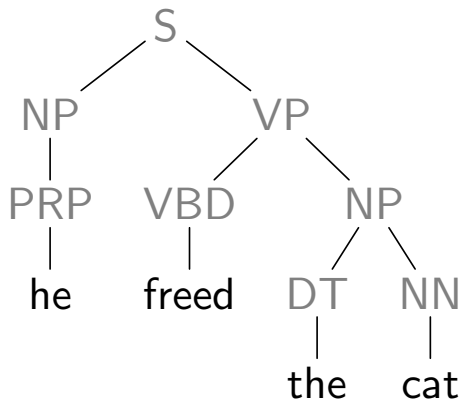


$\mapsto$ 0

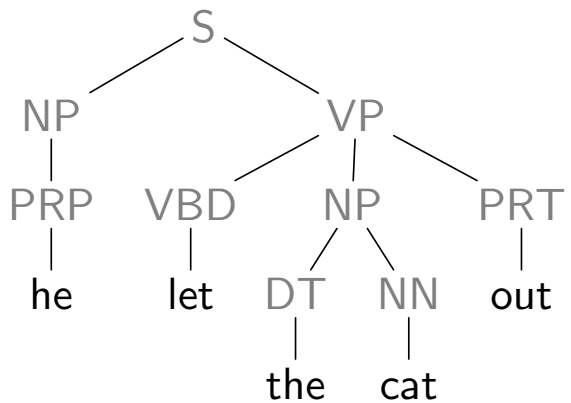
$\ldots$ $\mapsto$ 0

Output Projection $\pi_2 \colon \mathcal{T} \to \mathcal{L}$

$\pi_2\big((\mathrm{Yd}^{-1}(1.f) \odot \varphi') \lhd \tau \rhd \mathrm{Yd}^{-1}(\varphi)\big)$:



$\mapsto$ 0.25



$\mapsto$ 0.05

$\cdots \mapsto 0$

Language Yield $\mathsf{Yd}\colon \mathcal{L} \to \mathcal{K}$

$\mathsf{Yd}(\pi_2((\mathsf{Yd}^{-1}(1.f) \odot \varphi') \lhd \tau \rhd \mathsf{Yd}^{-1}(\varphi)))$:

$$\begin{aligned} \text{he freed the cat} &\mapsto 0.25 \\ \text{he let the cat out} &\mapsto 0.05 \\ \ldots &\mapsto 0 \end{aligned}$$

Best-Index Operation best: $\mathcal{K} \to \Sigma^*$

$\mathsf{best}(\mathsf{Yd}(\pi_2((\mathsf{Yd}^{-1}(1.f) \odot \varphi') \lhd \tau \rhd \mathsf{Yd}^{-1}(\varphi))))$:

he freed the cat

Effectiveness through Finite Representation (cf. Table 1.2)

closure property	references	complexity
$1.(\Sigma^*) \subseteq \mathcal{K}_{\mathrm{Rec}}$	(Berstel und Reutenauer 1988; Schützenberger 1961)	$O(n)$
$\mathrm{Yd}(\mathcal{L}_{\mathrm{Rec}}) \subseteq \mathcal{K}_{\mathrm{CF}}$	(Thatcher 1967; Ésik und Kuich 2003)	$O(r)$
$\mathrm{Yd}^{-1}(\mathcal{K}_{\mathrm{Rec}}) \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Maletti und Satta 2009)	$O(p^k)$
$\mathcal{L}_{\mathrm{Rec}} \odot \mathcal{L}_{\mathrm{Rec}} \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Borchardt 2004, Cor. 3.9)	$O(r_1 \cdot r_2)$
$\mathcal{L}_{\mathrm{Rec}} \lhd \mathcal{T}_{\mathrm{STSG}} \subseteq \mathcal{T}_{\mathrm{STSG}}$	(Maletti 2010)	$O(r_2 \cdot p_1^{k_2})$
$\mathcal{T}_{\mathrm{STSG}} \rhd \mathcal{L}_{\mathrm{Rec}} \subseteq \mathcal{T}_{\mathrm{STSG}}$	(Maletti 2010)	$O(r_1 \cdot p_2^{k_1})$
$\pi_2(\mathcal{T}_{\mathrm{STSG}}) \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Fülöp u. a. 2011)	$O(r)$
$\mathrm{best}(\mathcal{K}_{\mathrm{CF}}) \subseteq \Sigma^*$	(Knuth 1977; Huang und Chiang 2005; Büchse u. a. 2010)	$O(r \cdot \log p)$

Effectiveness through Finite Representation (cf. Table 1.2)

closure property	references	complexity
$1.(\Sigma^*) \subseteq \mathcal{K}_{\mathrm{Rec}}$	(Berstel und Reutenauer 1988; Schützenberger 1961)	$O(n)$
$\mathrm{Yd}(\mathcal{L}_{\mathrm{Rec}}) \subseteq \mathcal{K}_{\mathrm{CF}}$	(Thatcher 1967; Ésik und Kuich 2003)	$O(r)$
$\mathrm{Yd}^{-1}(\mathcal{K}_{\mathrm{Rec}}) \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Maletti und Satta 2009)	$O(p^k)$
$\mathcal{L}_{\mathrm{Rec}} \odot \mathcal{L}_{\mathrm{Rec}} \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Borchardt 2004, Cor. 3.9)	$O(r_1 \cdot r_2)$
$\mathcal{L}_{\mathrm{Rec}} \lhd \mathcal{T}_{\mathrm{STSG}} \subseteq \mathcal{T}_{\mathrm{STSG}}$	(Maletti 2010)	$O(r_2 \cdot p_1^{k_2})$
$\mathcal{T}_{\mathrm{STSG}} \rhd \mathcal{L}_{\mathrm{Rec}} \subseteq \mathcal{T}_{\mathrm{STSG}}$	(Maletti 2010)	$O(r_1 \cdot p_2^{k_1})$
$\pi_2(\mathcal{T}_{\mathrm{STSG}}) \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Fülöp u. a. 2011)	$O(r)$
$\mathrm{best}(\mathcal{K}_{\mathrm{CF}}) \subseteq \Sigma^*$	(Knuth 1977; Huang und Chiang 2005; Büchse u. a. 2010)	$O(r \cdot \log p)$

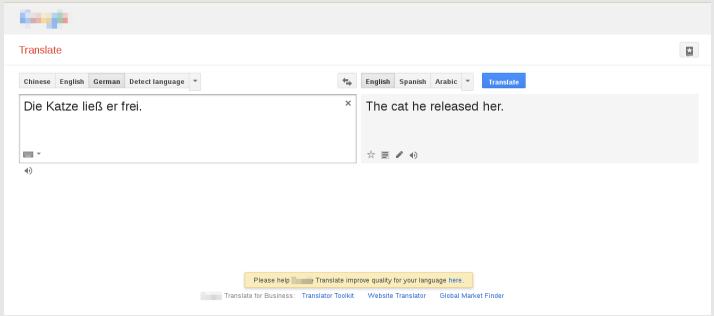
Effectiveness through Finite Representation (cf. Table 1.2)

closure property	references	complexity
$1.(\Sigma^*) \subseteq \mathcal{K}_{\mathrm{Rec}}$	(Berstel und Reutenauer 1988; Schützenberger 1961)	$O(n)$
$\mathrm{Yd}(\mathcal{L}_{\mathrm{Rec}}) \subseteq \mathcal{K}_{\mathrm{CF}}$	(Thatcher 1967; Ésik und Kuich 2003)	$O(r)$
$\mathrm{Yd}^{-1}(\mathcal{K}_{\mathrm{Rec}}) \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Maletti und Satta 2009)	$O(p^k)$
$\mathcal{L}_{\mathrm{Rec}} \odot \mathcal{L}_{\mathrm{Rec}} \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Borchardt 2004, Cor. 3.9)	$O(r_1 \cdot r_2)$
$\mathcal{L}_{\mathrm{Rec}} \lhd \mathcal{T}_{\mathrm{STSG}} \subseteq \mathcal{T}_{\mathrm{STSG}}$	(Maletti 2010)	$O(r_2 \cdot p_1^{k_2})$
$\mathcal{T}_{\mathrm{STSG}} \rhd \mathcal{L}_{\mathrm{Rec}} \subseteq \mathcal{T}_{\mathrm{STSG}}$	(Maletti 2010)	$O(r_1 \cdot p_2^{k_1})$
$\pi_2(\mathcal{T}_{\mathrm{STSG}}) \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Fülöp u. a. 2011)	$O(r)$
$\mathrm{best}(\mathcal{K}_{\mathrm{CF}}) \subseteq \Sigma^*$	(Knuth 1977; Huang und Chiang 2005; Büchse u. a. 2010)	$O(r \cdot \log p)$

coupling Statistical Machine Translation and Formal-Language Theory!

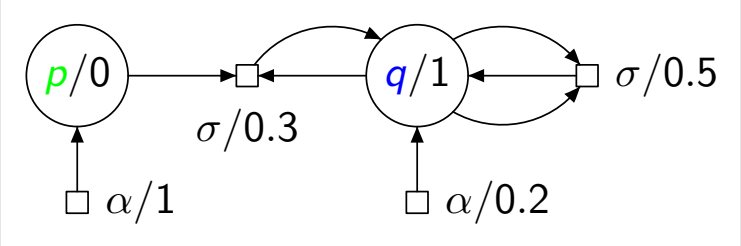
Overview

Statistical Machine Translation



✓ ✓

Formal-Language Theory



3

Decoder specification

algebraic decoder specification

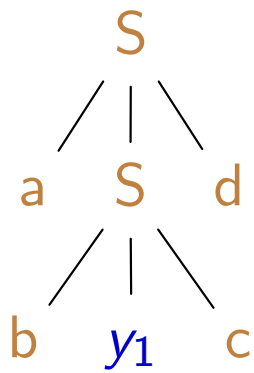
closure properties

binarization

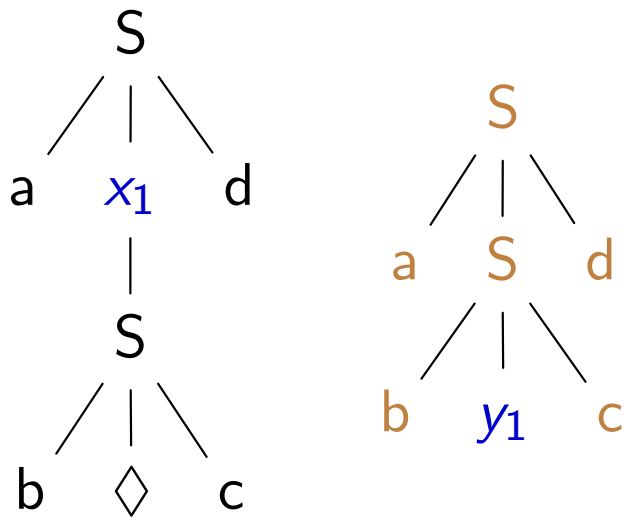
determinization

Outline

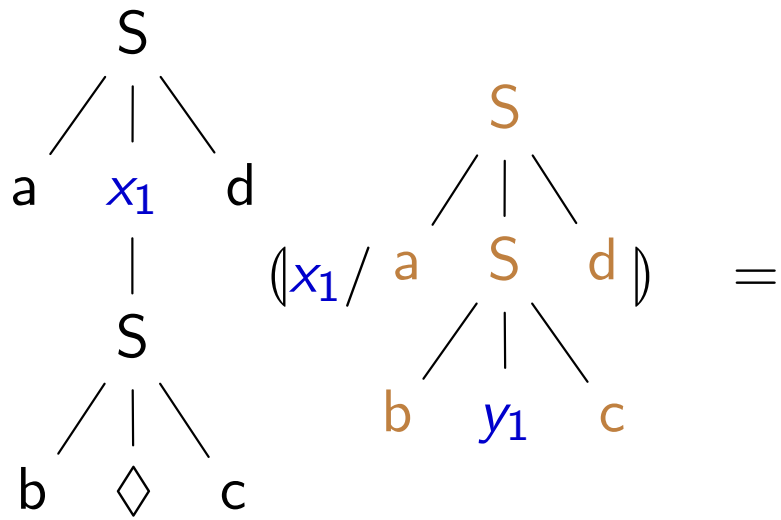
Second-Order Substitution



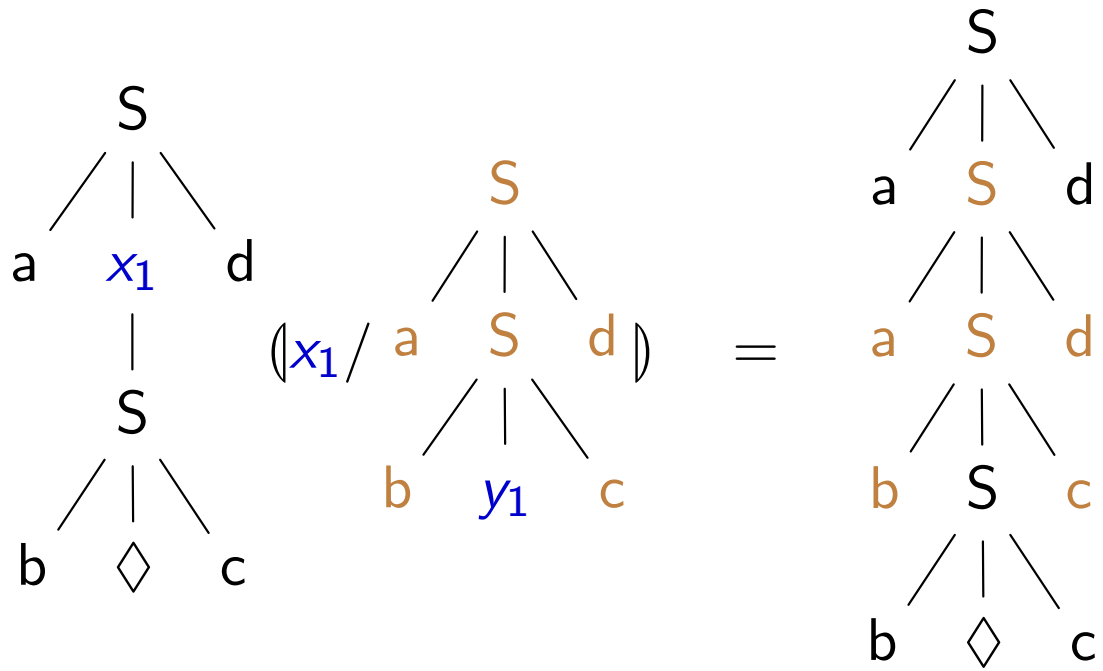
Second-Order Substitution



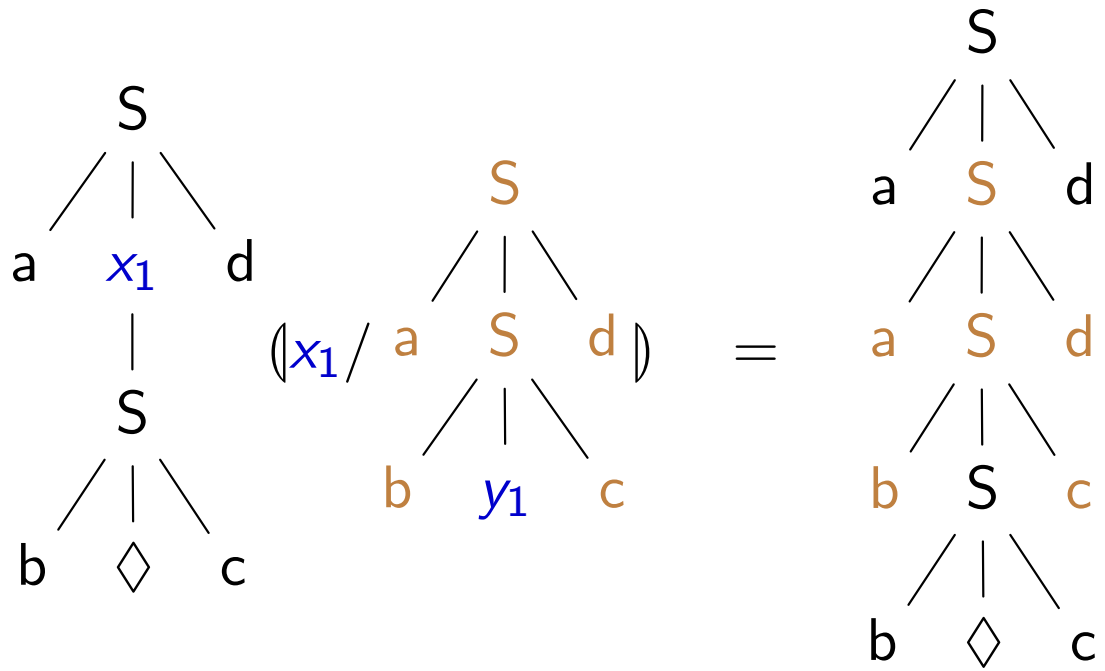
Second-Order Substitution



Second-Order Substitution



Second-Order Substitution



essential for describing real-world phenomena

(Shieber 2007; Kallmeyer u. a. 2009; Gildea 2010; Kaeshammer 2013)

Weighted Synchronous Context-Free Tree Grammar (WSCFTG)

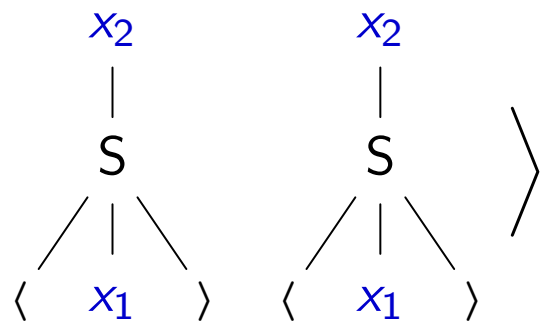
$$\mathrm{rk}(q) = 0, \quad \mathrm{rk}(f) = 1$$

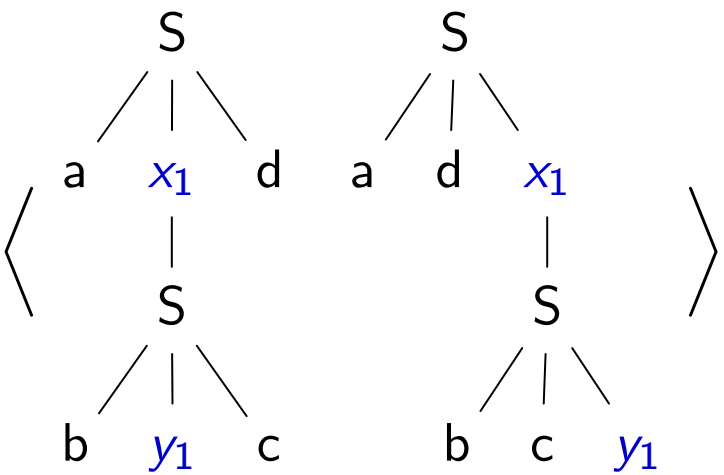
ρ_1: $q \to \alpha_1(q, f)$ # 0.3

ρ_2: $q \to \alpha_2()$ # 1

ρ_3: $f \to \alpha_3(f)$ # 0.7

ρ_4: $f \to \alpha_4()$ # 1

$\alpha_1 =$ 

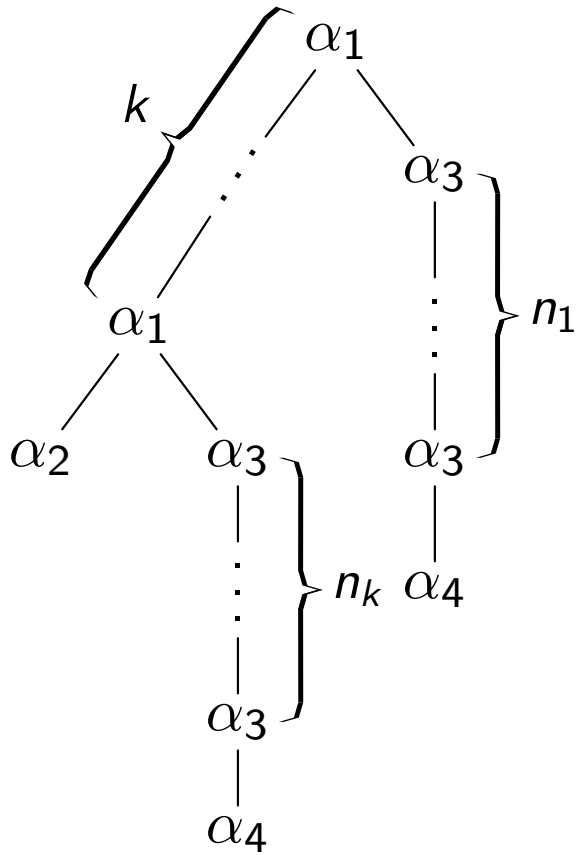
$\alpha_3 =$ 

$\alpha_2 = \langle \ \Diamond \quad \Diamond \ \rangle$

$\alpha_4 = \langle \ y_1 \quad y_1 \ \rangle$

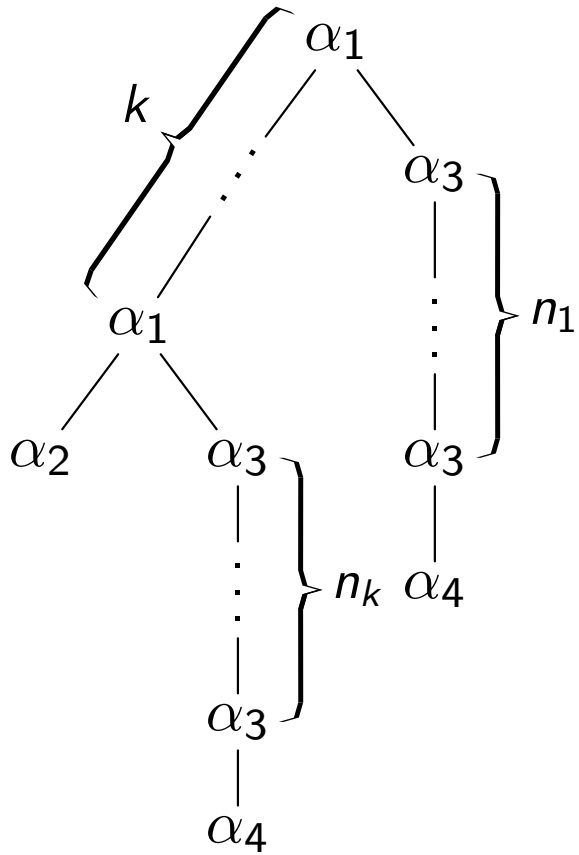
Center Trees and Derived Trees

(Arnold und Dauchet 1976)



Center Trees and Derived Trees

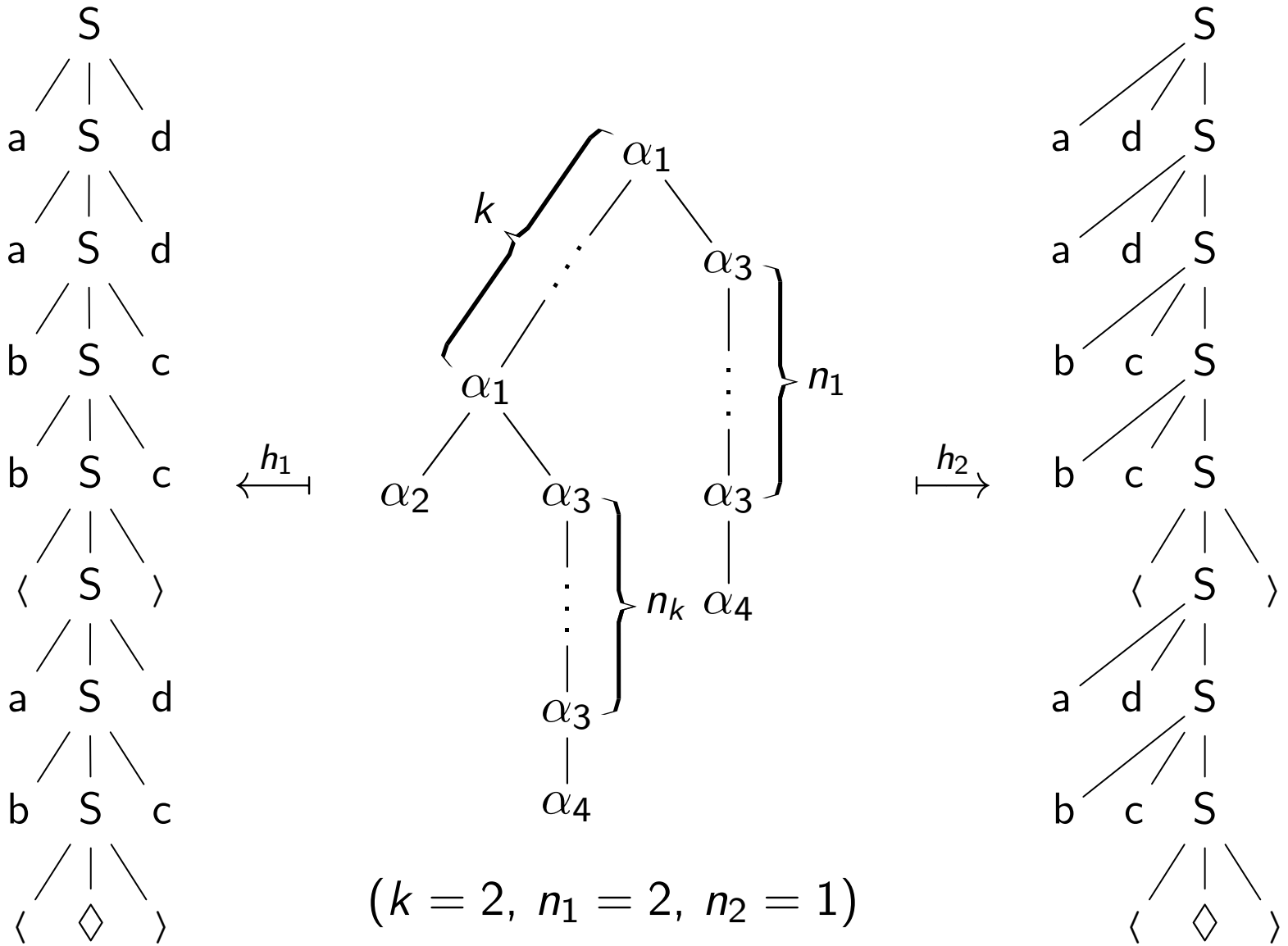
(Arnold und Dauchet 1976)



$(k = 2,\ n_1 = 2,\ n_2 = 1)$

Center Trees and Derived Trees

(Arnold und Dauchet 1976)



$(k = 2,\ n_1 = 2,\ n_2 = 1)$

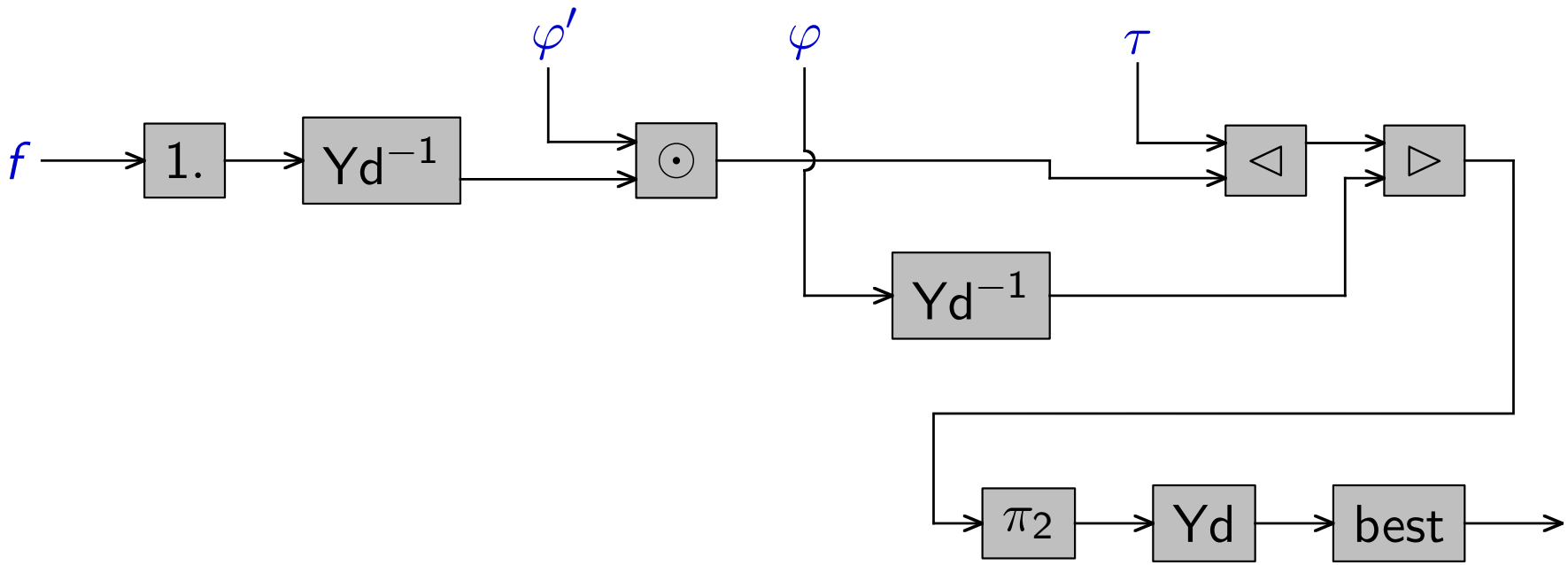
Subclasses of WSCFTG for Decoders

- Synchronous Tree-Substitution Grammar (STSG) (Eisner 2003),
- Synchronous Tree-Adjoining Grammar (STAG) (Shieber und Schabes 1990; Abeillé u. a. 1990),
- Synchronous Tree-Insertion Grammar (STIG) (Schabes und Waters 1994; Nesson 2009; DeNeefe 2011).

WSCFTG-based Decoder

$\Omega = \mathcal{T}_{\mathsf{WSCFTG}} \times \mathcal{K}_{\mathsf{Rec}} \times \mathcal{L}_{\mathsf{Rec}}$, $\quad \mathcal{S} = (\mathbb{R}_{\geq 0}, \max, \cdot, 0, 1)$

$\mathbb{D}(\tau, \varphi, \varphi') \colon \quad f \mapsto \mathsf{best}(\mathsf{Yd}(\pi_2((\mathsf{Yd}^{-1}(1.f) \odot \varphi') \lhd \tau \rhd \mathsf{Yd}^{-1}(\varphi))))$

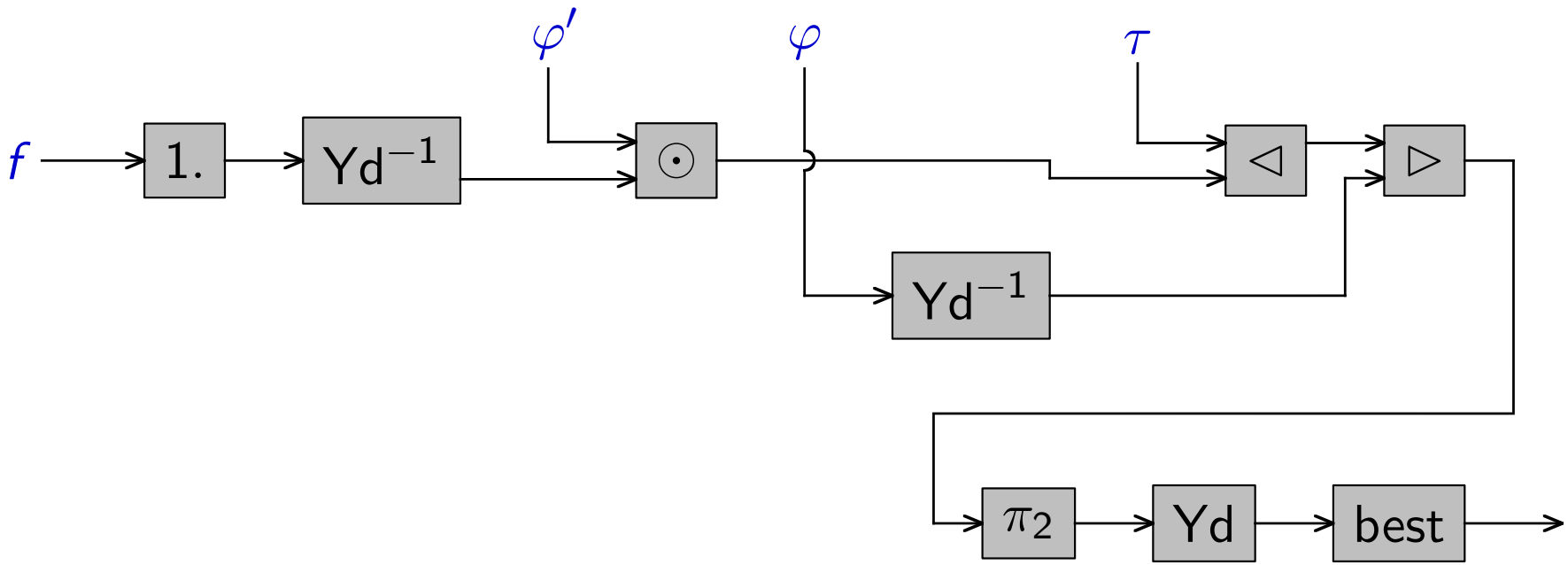


Effective? Extend table!

WSCFTG-based Decoder

$\Omega = \mathcal{T}_{\mathsf{WSCFTG}} \times \mathcal{K}_{\mathsf{Rec}} \times \mathcal{L}_{\mathsf{Rec}}$, $\quad \mathcal{S} = (\mathbb{R}_{\geq 0}, \max, \cdot, 0, 1)$

$\mathbb{D}(\tau, \varphi, \varphi') \colon \quad f \mapsto \mathsf{best}(\mathsf{Yd}(\pi_2((\mathsf{Yd}^{-1}(1.f) \odot \varphi') \lhd \tau \rhd \mathsf{Yd}^{-1}(\varphi))))$



Effective? Extend table! Prove the inclusions

$\mathcal{L}_{\mathsf{Rec}} \lhd \mathcal{T}_{\mathsf{WSCFTG}} \subseteq \mathcal{T}_{\mathsf{WSCFTG}}$, $\quad \mathcal{T}_{\mathsf{WSCFTG}} \rhd \mathcal{L}_{\mathsf{Rec}} \subseteq \mathcal{T}_{\mathsf{WSCFTG}}$,

maybe under suitable conditions.

Results I: Closure Property

Theorem 3.3.3

For commutative $\mathcal{S}$ we have

$$\mathcal{L}_{\mathsf{Rec}} \lhd \mathcal{T}_{\mathsf{WSCFTG}} \subseteq \mathcal{T}_{\mathsf{WSCFTG}}\,, \quad \mathcal{T}_{\mathsf{WSCFTG}} \rhd \mathcal{L}_{\mathsf{Rec}} \subseteq \mathcal{T}_{\mathsf{WSCFTG}}\,.$$

Results I: Closure Property

Theorem 3.3.3

For commutative $\mathcal{S}$ we have

$$\mathcal{L}_{\mathsf{Rec}} \lhd \mathcal{T}_{\mathsf{WSCFTG}} \subseteq \mathcal{T}_{\mathsf{WSCFTG}}\,, \quad \mathcal{T}_{\mathsf{WSCFTG}} \rhd \mathcal{L}_{\mathsf{Rec}} \subseteq \mathcal{T}_{\mathsf{WSCFTG}}\,.$$

Lemma 3.3.2 *There is effectively an admissible WSCFTG $M \lhd G = (Q', R', \mu', \nu')$ over Σ and $\mathcal{S}$ such that $M \lhd G$ is also a WTA over Γ and $\mathcal{S}$, $Q' = \bigcup_m Q^{(m)} \times P \times P^m$ with the ranks carried over from Q, $\nu'_{(q,p,\varepsilon)} = \nu_q \cdot (\nu_M)_p$, and the following holds. Let $\xi \in T_\Gamma$ be type conformant, $s = h_1(\xi)$, and $s \in C_\Sigma(m)$. Then there are families $(\equiv_{(p,p')} \mid p \in P, p' \in P^m)$ and $(\pi_{q'} \mid q' \in Q'^{(m)})$ such that*

- *$\equiv_{(p,p')}$ is an equivalence relation on $D^{(p,p')}(M, s)$,*
- *$\pi_{(q,p,p')} \colon D^{(q,p,p')}(M \lhd G, \xi) \to D^q(G, \xi) \times D^{(p,p')}(M, s)/_{\equiv_{(p,p')}}$ is bijective,*
- *$\pi_{q'}(d') = (d, D)$ implies $\langle d' \rangle = \langle d \rangle \cdot \sum_{e \in D} \langle e \rangle$.*

Results I: Closure Property

Theorem 3.3.3

For commutative $\mathcal{S}$ we have

$$\mathcal{L}_{\mathsf{Rec}} \lhd \mathcal{T}_{\mathsf{WSCFTG}} \subseteq \mathcal{T}_{\mathsf{WSCFTG}}, \quad \mathcal{T}_{\mathsf{WSCFTG}} \rhd \mathcal{L}_{\mathsf{Rec}} \subseteq \mathcal{T}_{\mathsf{WSCFTG}}.$$

Lemma 3.3.2 *There is effectively an admissible WSCFTG $M \lhd G = (Q', R', \mu', \nu')$ over Σ and $\mathcal{S}$ such that $M \lhd G$ is also a WTA over Γ and $\mathcal{S}$, $Q' = \bigcup_m Q^{(m)} \times P \times P^m$ with the ranks carried over from Q, $\nu'_{(q,p,\varepsilon)} = \nu_q \cdot (\nu_M)_p$, and the following holds. Let $\xi \in T_\Gamma$ be type conformant, $s = h_1(\xi)$, and $s \in C_\Sigma(m)$. Then there are families $(\equiv_{(p,p')} \mid p \in P, p' \in P^m)$ and $(\pi_{q'} \mid q' \in Q'^{(m)})$ such that*

- $\equiv_{(p,p')}$ *is an equivalence relation on* $D^{(p,p')}(M, s)$,
- $\pi_{(q,p,p')} \colon D^{(q,p,p')}(M \lhd G, \xi) \to D^q(G, \xi) \times D^{(p,p')}(M, s)/_{\equiv_{(p,p')}}$ *is bijective,*
- $\pi_{q'}(d') = (d, D)$ *implies* $\langle d' \rangle = \langle d \rangle \cdot \sum_{e \in D} \langle e \rangle$.

Time complexity of the construction: $O(|R| \cdot |P|^C)$ with
$C = \max\{\mathsf{rk}(q_0) + \cdots + \mathsf{rk}(q_l) + l + 1 \mid (q_1 \cdots q_l, \alpha, q_0) \in R\}$.

Results II: Algorithm

Algorithm 3.1

Algorithm based on Earley parsing (Earley 1970); computes $M \lhd G$, avoiding useless rules.

Results II: Algorithm

Algorithm 3.1

Algorithm based on Earley parsing (Earley 1970); computes $M \lhd G$, avoiding useless rules.

(1) $$\frac{}{(q_0, p_0)}$$

(2) $$\frac{(q, p)}{(\rho, \varepsilon, p)}$$

(3) $$\frac{(q, p) \qquad [\rho, \varepsilon, p, \theta]}{[q, p, p']} \;\Big\{\; p' = (\theta(y_1), \ldots, \theta(y_m))$$

(4) $$\frac{(\rho, w, p)}{[\rho, w, 0, p, p']} \;\Big\{\; (p', \zeta(w), p) \in R_M^{(\mathrm{rk}_\zeta(w))}$$

(5) $$\frac{[\rho, w, j, p, (p_1, \ldots, p_k)]}{(\rho, w(j+1), p_{j+1})} \;\Big\{\; 0 \le j < k$$

(6) $$\frac{[\rho, w, j, p, (p_1, \ldots, p_k)] \qquad [\rho, w(j+1), p_{j+1}]}{[\rho, w, j+1, p, (p_1, \ldots, p_k)]} \;\Big\{\; 0 \le j < k$$

(7) $$\frac{[\rho, w, p, \theta]}{[\rho, w, p]}$$

(8) $$\frac{(\rho, w, p)}{(q_\iota, p)} \;\Big\{\; \zeta(w) = x_\iota$$

(9) $$\frac{(\rho, w, p) \qquad [q_\iota, p, p']}{[\rho, w, 0, p, p']} \;\Big\{\; \zeta(w) = x_\iota$$

(10) $$\frac{(\rho, w, p)}{[\rho, w, p, \{y_\iota \mapsto p\}]} \;\Big\{\; \zeta(w) = y_\iota$$

(11) $$\frac{[\rho, w, k, p, p'] \qquad [\rho, w1, p_1, \theta_1] \quad \cdots \quad [\rho, wk, p_k, \theta_k]}{[\rho, w, p, \theta \cup \theta_1 \cup \cdots \cup \theta_k]} \;\Big\{\; p' = (p_1, \ldots, p_k)$$

where $\theta = \{\zeta(w) \mapsto (p, p')\}$ if $\zeta(w) \in X$, and
$\theta = \emptyset$ otherwise

Note: we assume that $\rho \in R$, $\rho = (q_1 \cdots q_l, \langle \zeta\zeta' \rangle, q)$, and $q \in Q^{(m)}$.

Results II: Algorithm

Algorithm 3.1

Algorithm based on Earley parsing (Earley 1970); computes $M \lhd G$, avoiding useless rules.

Theorem 3.4.7

The algorithm is correct and complete.

(1) $$\frac{}{(q_0, p_0)}$$

(2) $$\frac{(q, p)}{(\rho, \varepsilon, p)}$$

(3) $$\frac{(q, p) \qquad [\rho, \varepsilon, p, \theta]}{[q, p, p']} \;\{\; p' = (\theta(y_1), \ldots, \theta(y_m))$$

(4) $$\frac{(\rho, w, p)}{[\rho, w, 0, p, p']} \;\Big\{\; (p', \zeta(w), p) \in R_M^{(\mathrm{rk}_\zeta(w))}$$

(5) $$\frac{[\rho, w, j, p, (p_1, \ldots, p_k)]}{(\rho, w(j+1), p_{j+1})} \;\{\; 0 \le j < k$$

(6) $$\frac{[\rho, w, j, p, (p_1, \ldots, p_k)] \qquad [\rho, w(j+1), p_{j+1}]}{[\rho, w, j+1, p, (p_1, \ldots, p_k)]} \;\{\; 0 \le j < k$$

(7) $$\frac{[\rho, w, p, \theta]}{[\rho, w, p]}$$

(8) $$\frac{(\rho, w, p)}{(q_\iota, p)} \;\{\; \zeta(w) = x_\iota$$

(9) $$\frac{(\rho, w, p) \qquad [q_\iota, p, p']}{[\rho, w, 0, p, p']} \;\{\; \zeta(w) = x_\iota$$

(10) $$\frac{(\rho, w, p)}{[\rho, w, p, \{y_\iota \mapsto p\}]} \;\{\; \zeta(w) = y_\iota$$

(11) $$\frac{[\rho, w, k, p, p'] \qquad \begin{matrix}[\rho, w1, p_1, \theta_1] \\ \cdots\ [\rho, wk, p_k, \theta_k]\end{matrix}}{[\rho, w, p, \theta \cup \theta_1 \cup \cdots \cup \theta_k]} \;\{\; p' = (p_1, \ldots, p_k)$$

where $\theta = \{\zeta(w) \mapsto (p, p')\}$ if $\zeta(w) \in X$, and
$\theta = \emptyset$ otherwise

Note: we assume that $\rho \in R$, $\rho = (q_1 \cdots q_l, \langle \zeta\zeta' \rangle, q)$, and $q \in Q^{(m)}$.

Results II: Algorithm

Algorithm 3.1

Algorithm based on Earley parsing (Earley 1970); computes $M \lhd G$, avoiding useless rules.

Theorem 3.4.7

The algorithm is correct and complete.

Time complexity of the algorithm:

$O(|G|_{\mathsf{in}} \cdot |R_M| \cdot |P|^C)$

(1) $\dfrac{}{(q_0, p_0)}$

(2) $\dfrac{(q,p)}{(\rho, \varepsilon, p)}$

(3) $\dfrac{(q,p) \qquad [\rho, \varepsilon, p, \theta]}{[q, p, p']} \;\{\; p' = (\theta(y_1), \ldots, \theta(y_m))$

(4) $\dfrac{(\rho, w, p)}{[\rho, w, 0, p, p']} \;\Big\{\; (p', \zeta(w), p) \in R_M^{(\mathrm{rk}_\zeta(w))}$

(5) $\dfrac{[\rho, w, j, p, (p_1, \ldots, p_k)]}{(\rho, w(j+1), p_{j+1})} \;\{\; 0 \le j < k$

(6) $\dfrac{[\rho, w, j, p, (p_1, \ldots, p_k)] \qquad [\rho, w(j+1), p_{j+1}]}{[\rho, w, j+1, p, (p_1, \ldots, p_k)]} \;\{\; 0 \le j < k$

(7) $\dfrac{[\rho, w, p, \theta]}{[\rho, w, p]}$

(8) $\dfrac{(\rho, w, p)}{(q_\iota, p)} \;\{\; \zeta(w) = x_\iota$

(9) $\dfrac{(\rho, w, p) \qquad [q_\iota, p, p']}{[\rho, w, 0, p, p']} \;\{\; \zeta(w) = x_\iota$

(10) $\dfrac{(\rho, w, p)}{[\rho, w, p, \{y_\iota \mapsto p\}]} \;\{\; \zeta(w) = y_\iota$

(11) $\dfrac{[\rho, w, k, p, p'] \qquad [\rho, w1, p_1, \theta_1] \quad \cdots \quad [\rho, wk, p_k, \theta_k]}{[\rho, w, p, \theta \cup \theta_1 \cup \cdots \cup \theta_k]} \;\{\; p' = (p_1, \ldots, p_k)$

where $\theta = \{\zeta(w) \mapsto (p, p')\}$ if $\zeta(w) \in X$, and
$\theta = \emptyset$ otherwise

Note: we assume that $\rho \in R$, $\rho = (q_1 \cdots q_l, \langle \zeta\zeta' \rangle, q)$, and $q \in Q^{(m)}$.

Outline

Effectiveness through Finite Representation (cf. Table 1.2)

closure property	references	complexity
$1.(\Sigma^*) \subseteq \mathcal{K}_{\mathrm{Rec}}$	(Berstel und Reutenauer 1988; Schützenberger 1961)	$O(n)$
$\mathrm{Yd}(\mathcal{L}_{\mathrm{Rec}}) \subseteq \mathcal{K}_{\mathrm{CF}}$	(Thatcher 1967; Ésik und Kuich 2003)	$O(r)$
$\mathrm{Yd}^{-1}(\mathcal{K}_{\mathrm{Rec}}) \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Maletti und Satta 2009)	$O(p^k)$
$\mathcal{L}_{\mathrm{Rec}} \odot \mathcal{L}_{\mathrm{Rec}} \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Borchardt 2004, Cor. 3.9)	$O(r_1 \cdot r_2)$
$\mathcal{L}_{\mathrm{Rec}} \lhd \mathcal{T}_{\mathrm{STSG}} \subseteq \mathcal{T}_{\mathrm{STSG}}$	(Maletti 2010)	$O(r_2 \cdot p_1^{k_2})$
$\mathcal{T}_{\mathrm{STSG}} \rhd \mathcal{L}_{\mathrm{Rec}} \subseteq \mathcal{T}_{\mathrm{STSG}}$	(Maletti 2010)	$O(r_1 \cdot p_2^{k_1})$
$\pi_2(\mathcal{T}_{\mathrm{STSG}}) \subseteq \mathcal{L}_{\mathrm{Rec}}$	(Fülöp u. a. 2011)	$O(r)$
$\mathrm{best}(\mathcal{K}_{\mathrm{CF}}) \subseteq \Sigma^*$	(Knuth 1977; Huang und Chiang 2005; Büchse u. a. 2010)	$O(r \cdot \log p)$

Concepts

Binarization Mapping

$\mathsf{bin} : \mathcal{F} \to \mathcal{F}$

- preserves meaning,
- "linear growth": $|R'| \leq k \cdot |R|$

Concepts

Binarization Mapping

$\mathsf{bin} : \mathcal{F} \to \mathcal{F}$

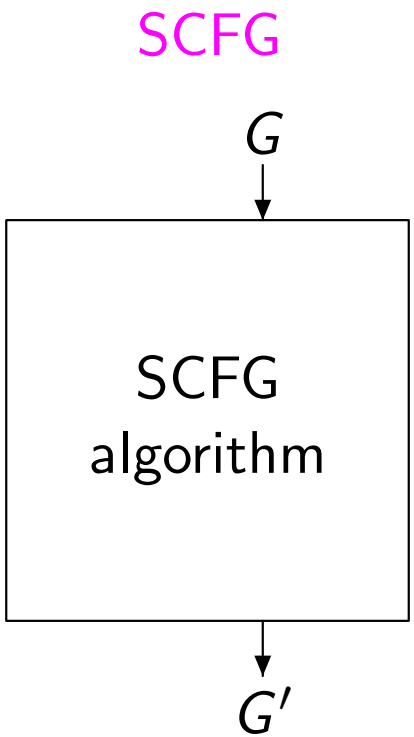
- preserves meaning,
- "linear growth": $|R'| \leq k \cdot |R|$

Binarization Domain

$\mathsf{bdom}(\mathsf{bin}) = \{G \mid G \in \mathcal{F}, \mathsf{rk}(\mathsf{bin}(G)) \leq 2\}$

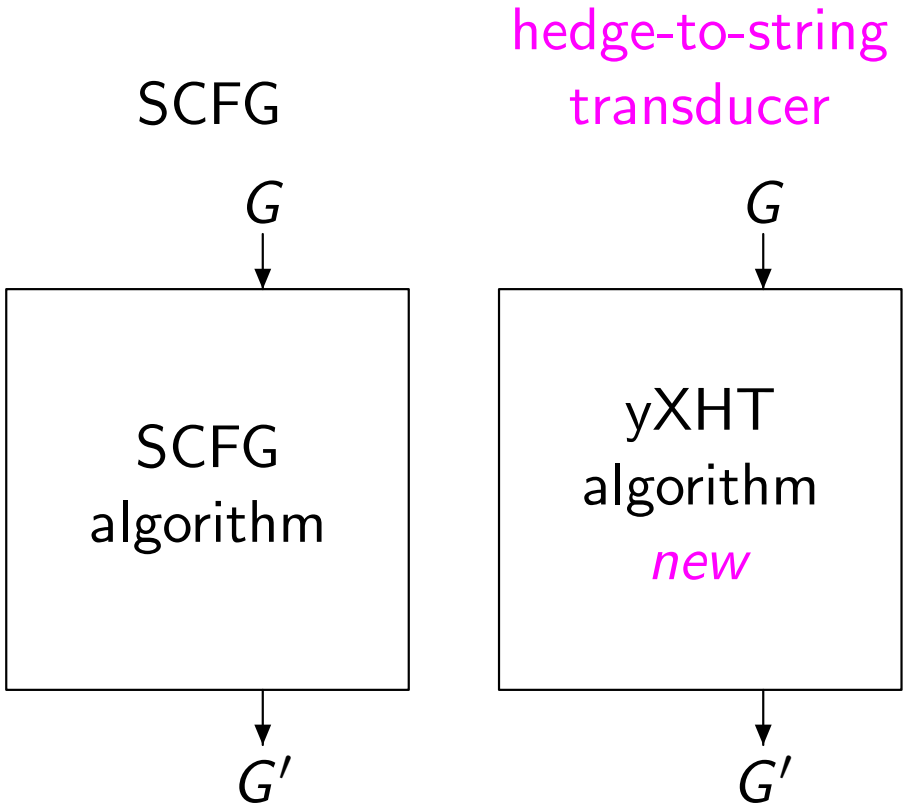
Result: Generic Algorithm + Instances

Rule-by-rule complete binarization mappings



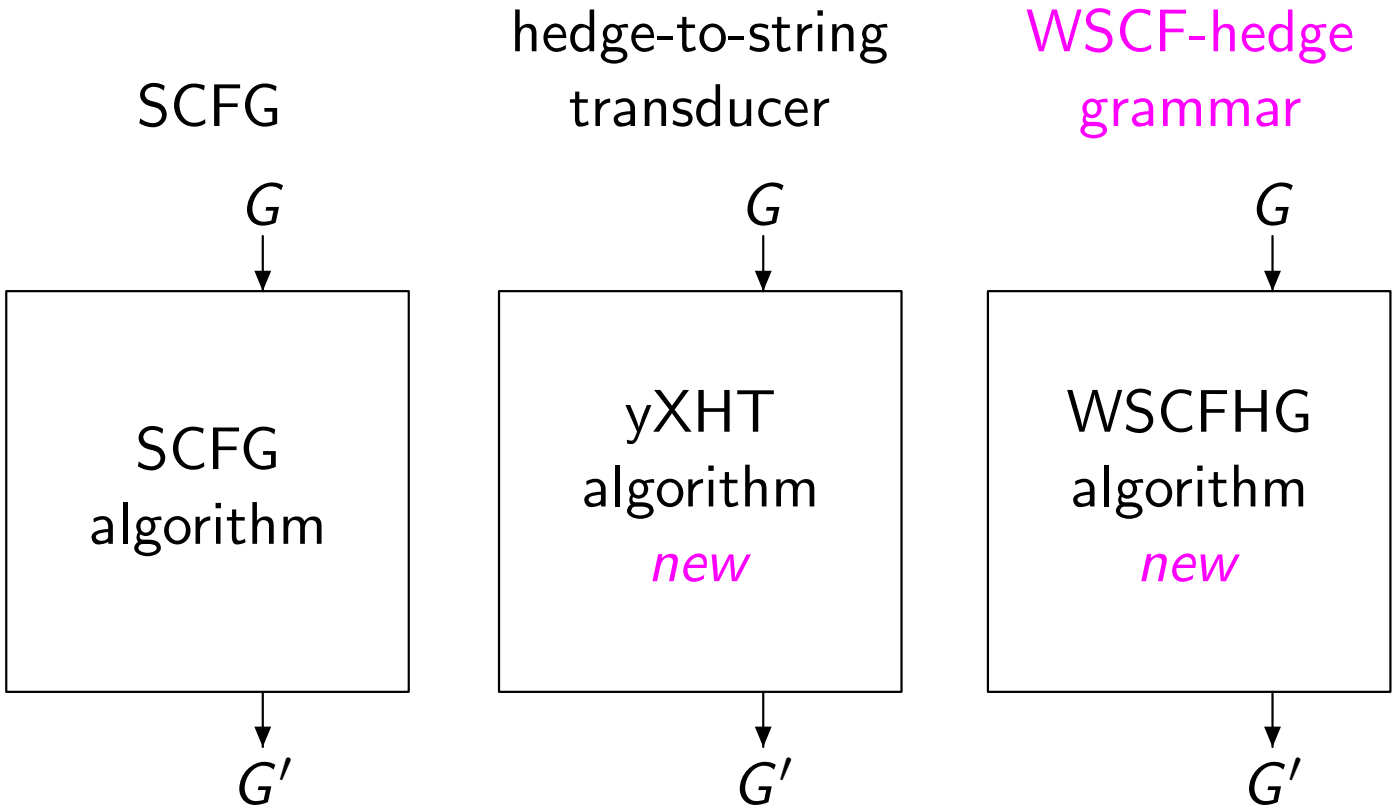
Result: Generic Algorithm + Instances

Rule-by-rule complete binarization mappings



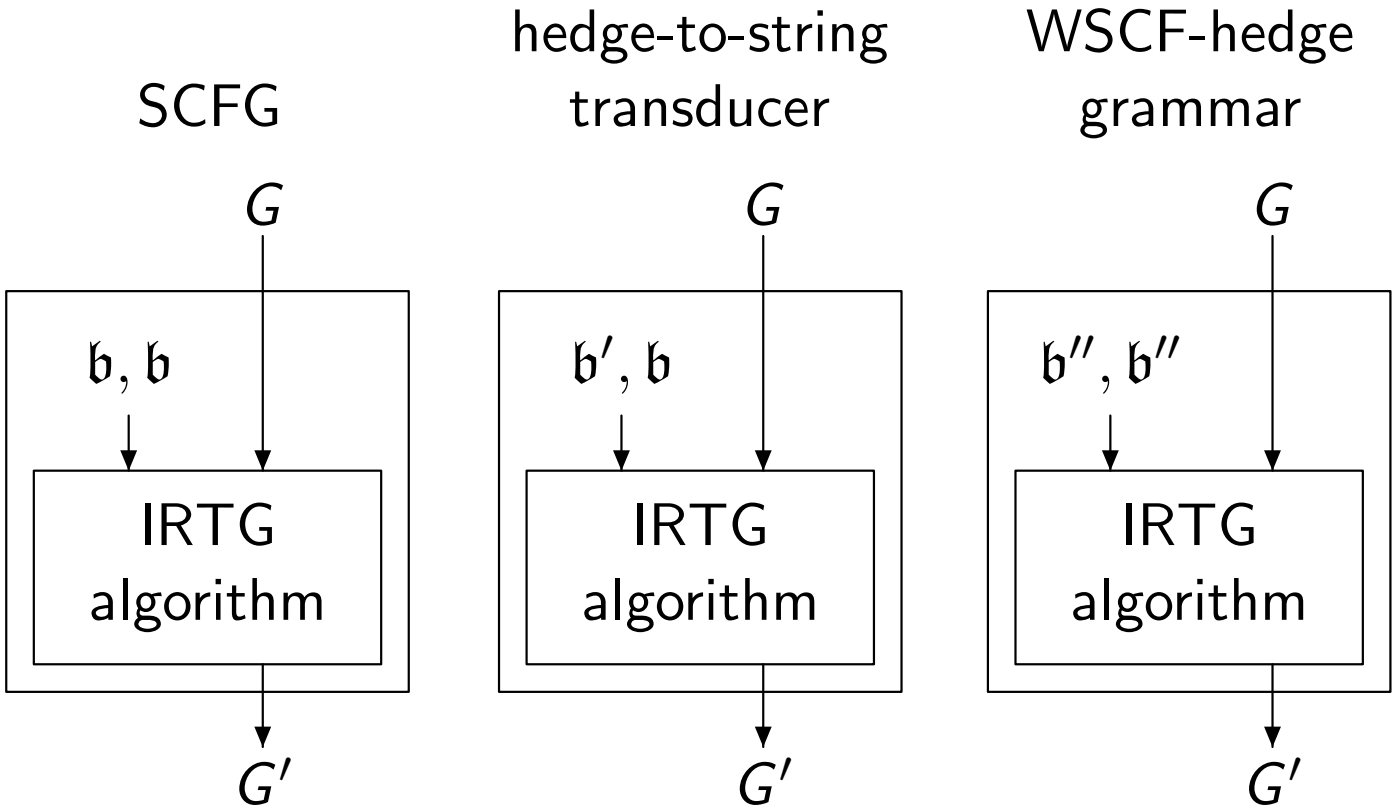
Result: Generic Algorithm + Instances

Rule-by-rule complete binarization mappings



Result: Generic Algorithm + Instances

Rule-by-rule complete binarization mappings



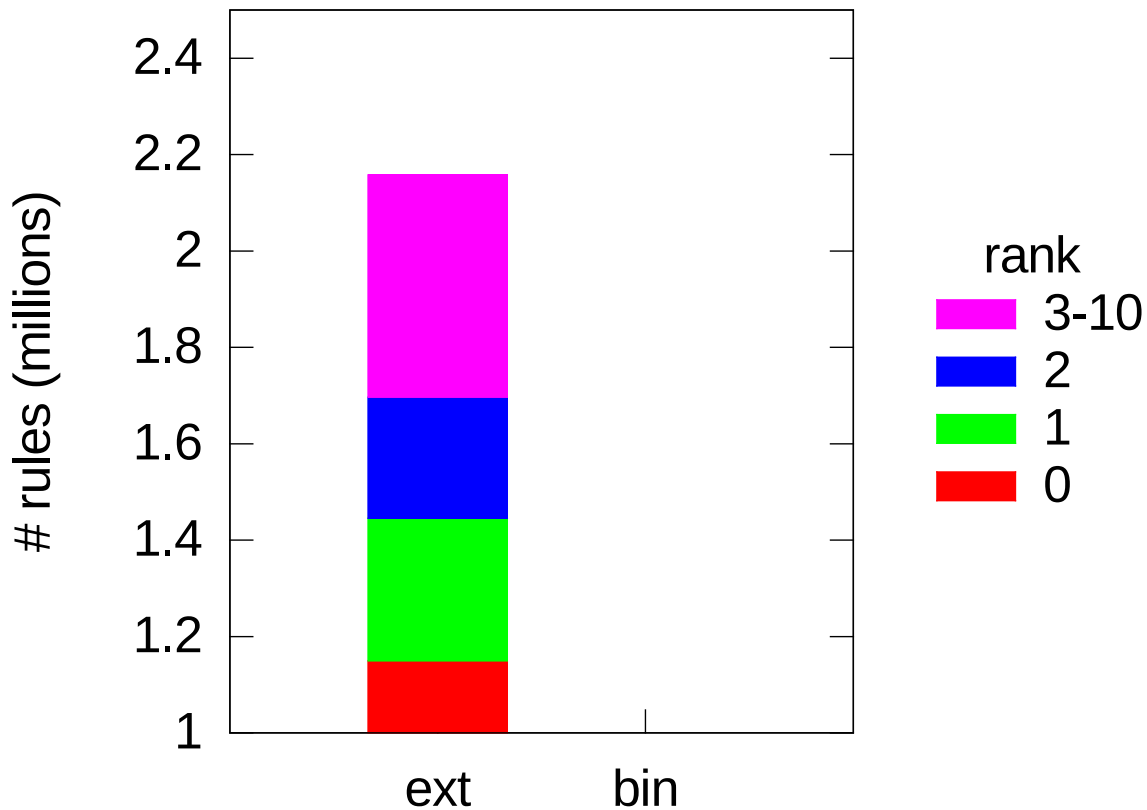
. . . with merely three b-rules $\mathfrak{b}, \mathfrak{b}', \mathfrak{b}''$!

Experiment: Hedge-to-String Transducer

- extracted from 1 million sentence pairs (Europarl Eng-Ger)

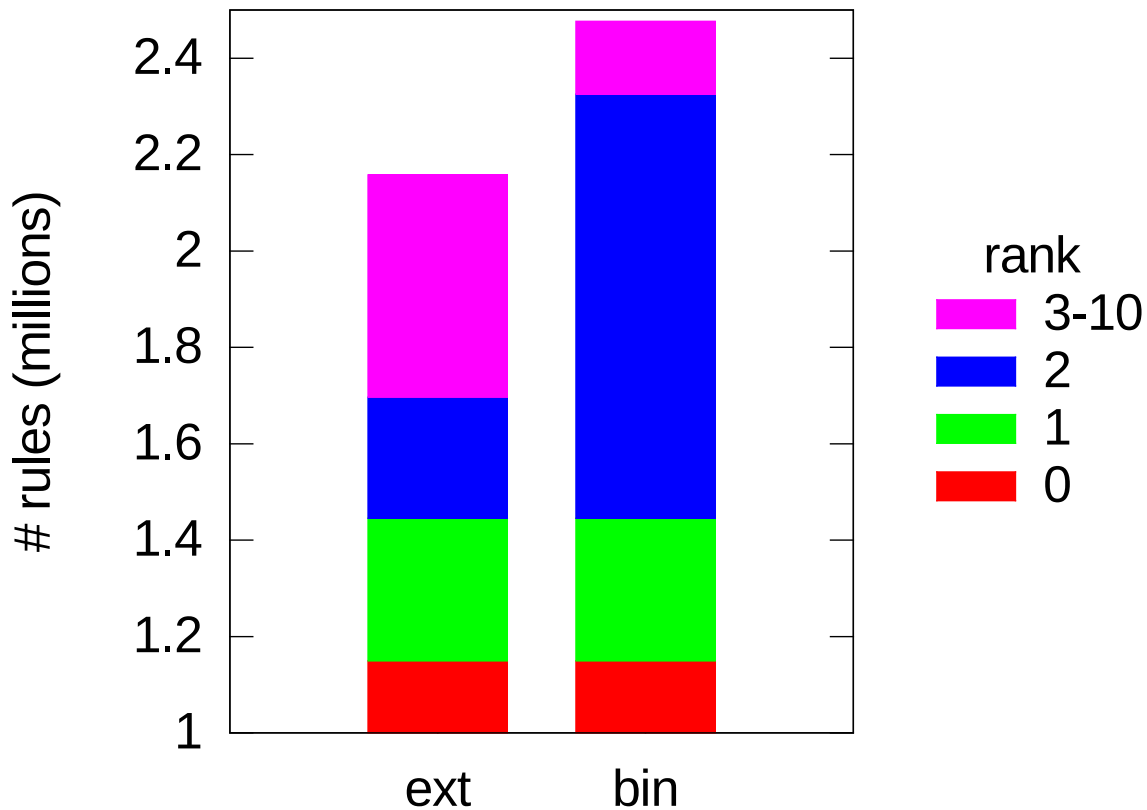
Experiment: Hedge-to-String Transducer

- extracted from 1 million sentence pairs (Europarl Eng-Ger)
- 2.15 million rules up to rank 10



Experiment: Hedge-to-String Transducer

- extracted from 1 million sentence pairs (Europarl Eng-Ger)
- 2.15 million rules up to rank 10
- 67 % could be processed



Outline

Results I: Determinization

formalism	restriction	semiring	restriction
literature			

Results I: Determinization

formalism	restriction	semiring restriction	
	literature		
FSA	–	Boolean	
WSA	twins property	tropical	
WSA	twins property	commutative, extremal	(1)

legend: (1) requires a maximal factorization from the user

Results I: Determinization

formalism	restriction	semiring restriction	
	literature		
FSA	–	Boolean	
WSA	twins property	tropical	
WSA	twins property	commutative, extremal	(1)
FTA	–	Boolean	
WTA	–	locally finite, semifield	
WTA	–	locally finite	
WTA	acyclic	nonnegative reals	(2)

legend: (1) requires a maximal factorization from the user
(2) lacks a proof

Results I: Determinization

formalism	restriction	semiring restriction	
	literature		
FSA	–	Boolean	
WSA	twins property	tropical	
WSA	twins property	commutative, extremal	(1)
FTA	–	Boolean	
WTA	–	locally finite, semifield	
WTA	–	locally finite	
WTA	acyclic	nonnegative reals	(2)
	Theorem 5.5.3		
WTA	twins property	commutative, extremal	(1)
WTA	–	locally finite	
WTA	acyclic	commutative	

legend: (1) requires a maximal factorization from the user
(2) lacks a proof

Results I: Determinization

formalism	restriction	semiring restriction	
	literature		
FSA	–	Boolean	
WSA	twins property	tropical	
WSA	twins property	commutative, extremal	(1)
FTA	–	Boolean	
WTA	–	locally finite, semifield	
WTA	–	locally finite	
WTA	acyclic	nonnegative reals	(2)
	Theorem 5.5.3		
WTA	twins property	commutative, extremal	(1)
WTA	–	locally finite	
WTA	acyclic	commutative	

legend: (1) requires a maximal factorization from the user
(2) lacks a proof

Results I: Determinization

formalism	restriction	semiring restriction	
	literature		
FSA	–	Boolean	
WSA	twins property	tropical	
WSA	twins property	commutative, extremal	(1)
FTA	–	Boolean	
WTA	–	locally finite, semifield	
WTA	–	locally finite	
WTA	acyclic	nonnegative reals	(2)
	Theorem 5.5.3		
WTA	twins property	commutative, extremal	(1)
WTA	–	locally finite	
WTA	acyclic	commutative	

legend: (1) requires a maximal factorization from the user
(2) lacks a proof

Results I: Determinization

formalism	restriction	semiring restriction	
	literature		
FSA	–	Boolean	
WSA	twins property	tropical	
WSA	twins property	commutative, extremal	(1)
FTA	–	Boolean	
WTA	–	locally finite, semifield	
WTA	–	locally finite	
WTA	acyclic	nonnegative reals	(2)
	Theorem 5.5.3		
WTA	twins property	commutative, extremal	(1)
WTA	–	locally finite	
WTA	acyclic	commutative	

legend: (1) requires a maximal factorization from the user
(2) lacks a proof

Results II: Decidability

Literature

The twins property is decidable for the classes

- of cycle-unambiguous WSA over commutative cancellative semirings,
- of WSA over the tropical semiring.

Results II: Decidability

Literature

The twins property is decidable for the classes

- of cycle-unambiguous WSA over commutative cancellative semirings,
- of WSA over the tropical semiring.

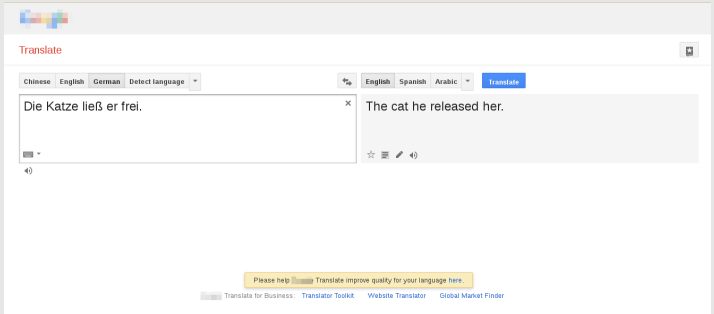
Theoreme 5.5.4, 5.5.5

The twins property is decidable for the classes

- of cycle-unambiguous WTA over commutative zero-sum-free zero-divisor-free semirings,
- of WTA over extremal semifields.

Summary

Statistical Machine Translation

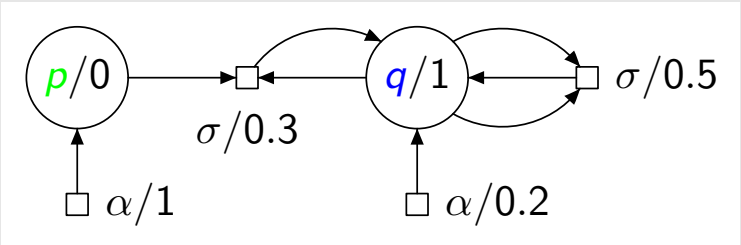


✓ ✓

Decoder specification

algebraic decoder specification

Formal-Language Theory



✓

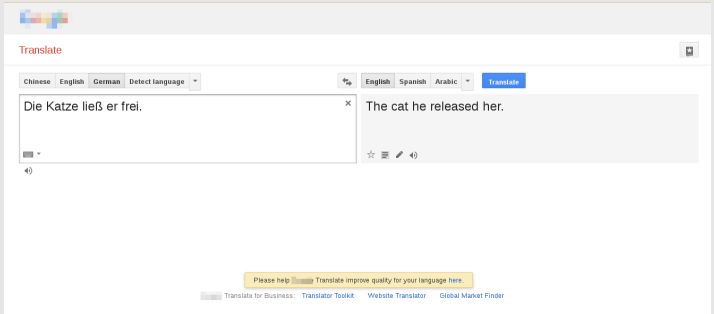
closure properties

binarization

determinization

Summary

Statistical Machine Translation

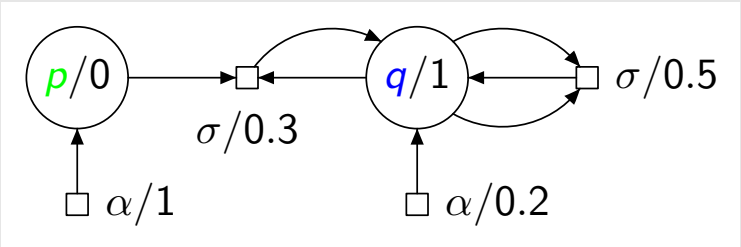


✓ ✓

Decoder specification

algebraic decoder specification

Formal-Language Theory



✓

closure properties

binarization

determinization

Thank you for your attention.

Bibliography I

- Abeillé, Anne, Yves Schabes und Aravind K. Joshi (1990). "Using lexicalized tags for machine translation". In: Proceedings of the Thirteenth International Conference on Computational Linguistics (COLING '90). Bd. 3, S. 1–6.
- Aho, Alfred V. und Jeffrey D. Ullman (1969). "Syntax directed translations and the pushdown assembler". In: Journal of Computer and System Sciences 3, S. 37–56.
- Alexandrakis, Athanasios und Symeon Bozapalidis (1987). "Weighted grammars and Kleene's theorem". In: Information Processing Letters 24.1, S. 1–4.
- Arnold, André und Max Dauchet (1976). "Bi-transduction de forêts". In: Proc. 3rd Int. Coll. Automata, Languages and Programming, S. 74–86.
- Berstel, Jean und Christophe Reutenauer (1988). Rational Series and Their Languages. Bd. 12. EATCS Monographs on Theoretical Computer Science. Springer.

Bibliography II


Borchardt, Björn (2004). "A pumping lemma and decidability problems for recognizable tree series". In: Acta Cybernet. 16.4, S. 509–544.


Büchse, Matthias, Daniel Geisler, Torsten Stüber und Heiko Vogler (2010). "n-Best Parsing Revisited". In: Proceedings of the 2010 Workshop on Applications of Tree Automata in Natural Language Processing, ACL 2010. Uppsala, Sweden, 16 July 2010, S. 46–54.


Chang, Yin-Wen und Michael Collins (2011). "Exact Decoding of Phrase-Based Translation Models through Lagrangian Relaxation". In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, S. 26–37. URL: http://www.aclweb.org/anthology/D11-1003.


Chiang, David (2007). "Hierarchical Phrase-Based Translation". In: Comp. Ling. 33.2, S. 201–228.

Bibliography III


Chiang, David, Adam Lopez, Nitin Madnani, Christof Monz, Philip Resnik und Michael Subotin (2005). "The Hiero machine translation system: extensions, evaluation, and analysis". In: HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, S. 779–786.


DeNeefe, Steve (2011). "Tree-Adjoining Machine Translation". Diss. University of Southern California.


Dyer, Chris, Adam Lopez, Juri Ganitkevitch, Jonathan Weese, Ferhan Ture, Phil Blunsom, Hendra Setiawan, Vladimir Eidelman und Philip Resnik (2010). "cdec: A Decoder, Alignment, and Learning Framework for Finite-State and Context-Free Translation Models". In: Proceedings of the ACL 2010 System Demonstrations, S. 7–12. URL: http://www.aclweb.org/anthology/P10-4002.

Bibliography IV

- Earley, Jay (1970). "An Efficient Context-Free Parsing Algorithm". In: Communications of the ACM 13.2, S. 94–102.
- Eisner, Jason (2003). "Learning non-isomorphic tree mappings for machine translation". In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 2. ACL '03, S. 205–208.
- Ésik, Zoltán und Werner Kuich (2003). "Formal Tree Series". In: J. Autom. Lang. Combin. 8.2, S. 219–285.
- Fülöp, Zoltán, Andreas Maletti und Heiko Vogler (2011). "Weighted Extended Tree Transducers". In: Fundam. Inform. 111.2, S. 163–202.
- Fülöp, Zoltán und Heiko Vogler (2009). "Weighted tree automata and tree transducers". In: Handbook of Weighted Automata. Hrsg. von Manfred Droste, Werner Kuich und Heiko Vogler. EATCS Monographs in Theoretical Computer Science. Kap. 9.

Bibliography V


Gildea, Daniel (2010). "Optimal Parsing Strategies for Linear Context-Free Rewriting Systems". In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, S. 769–776. URL: http://www.aclweb.org/anthology/N10-1118.


Goguen, Joseph A., Jim W. Thatcher, Eric G. Wagner und Jesse B. Wright (1977). "Initial algebra semantics and continuous algebras". In: J. ACM 24, S. 68–95.


Golan, Jonathan S. (1999). Semirings and their Applications. Kluwer Academic.


Hebisch, Udo und Hanns Joachim Weinert (1998). Semirings: Algebraic Theory and Applications in Computer Science. Bd. 5. Series in Algebra. World Scientific.

Bibliography VI


Huang, Liang und David Chiang (2005). "Better k-best parsing". In: Parsing '05: Proceedings of the Ninth International Workshop on Parsing Technology, S. 53–64. URL: http://www.cis.upenn.edu/~lhuang3/huang-iwpt-correct.pdf.


Huang, Liang, Hao Zhang, Daniel Gildea und Kevin Knight (2009). "Binarization of Synchronous Context-Free Grammars". In: Comp. Ling. 35.4, S. 559–595. URL: http://www.aclweb.org/anthology/J/J09/J09-4009.pdf.


Kaeshammer, Miriam (2013). "Synchronous Linear Context-Free Rewriting Systems for Machine Translation". In: Proceedings of the Seventh Workshop on Syntax, Semantics and Structure in Statistical Translation, S. 68–77. URL: http://www.aclweb.org/anthology/W13-0808.

Bibliography VII


Kallmeyer, Laura, Wolfgang Maier und Giorgio Satta (2009). "Synchronous Rewriting in Treebanks". In: Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09), S. 69–72. URL: http://www.aclweb.org/anthology/W09-3810.


Knuth, Donald E. (1977). "A Generalization of Dijkstra's Algorithm". In: Inform. Process. Lett. 6.1, S. 1–5.


Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin und Evan Herbst (2007). "Moses: open source toolkit for statistical machine translation". In: Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions. ACL '07, S. 177–180. URL: http://www.aclweb.org/anthology/P07-2045.

Bibliography VIII


Koller, Alexander und Marco Kuhlmann (2011). "A Generalized View on Parsing and Translation". In: Proceedings of the 12th International Conference on Parsing Technologies, S. 2–13. URL: http://www.aclweb.org/anthology/W11-2902.


Kuich, Werner (1998). "Formal power series over trees". In: 3rd International Conference on Developments in Language Theory, DLT 1997, Thessaloniki, Greece, Proceedings. Hrsg. von Symeon Bozapalidis, S. 61–101.


Lewis, Philip M. und Richard E. Stearns (1966). "Syntax directed transduction". In: Foundations of Computer Science, IEEE Annual Symposium on, S. 21–35.

Bibliography IX


Li, Zhifei, Chris Callison-Burch, Chris Dyer, Juri Ganitkevitch, Sanjeev Khudanpur, Lane Schwartz, Wren N. G. Thornton, Jonathan Weese und Omar F. Zaidan (2009). "Joshua: an open source toolkit for parsing-based machine translation". In: Proceedings of the Fourth Workshop on Statistical Machine Translation. StatMT '09, S. 135–139. URL: http://dl.acm.org/citation.cfm?id=1626431.1626459.


Maletti, Andreas (2010). "Input and Output Products for Weighted Extended Top-down Tree Transducers". In: Proc. 14th Int. Conf. Developments in Language Theory. Hrsg. von Yuan Gao, Hanlin Lu, Shinnosuke Seki und Sheng Yu. Bd. 6224. LNCS, S. 316–327.


Maletti, Andreas und Giorgio Satta (2009). "Parsing Algorithms based on Tree Automata". In: Proc. 11th Int. Conf. Parsing Technologies, S. 1–12.

Bibliography X

- May, Jonathan und Kevin Knight (2006). "A better N-best list: practical determinization of weighted finite tree automata". In: Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, S. 351–358.
- Nesson, Rebecca (2009). "Synchronous and Multicomponent Tree-Adjoining Grammars: Complexity, Algorithms and Linguistic Applications". Diss. Harvard University.
- Rush, Alexander M. und Michael Collins (2011). "Exact Decoding of Syntactic Translation Models through Lagrangian Relaxation". In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Techologies, S. 72–82. URL: http://www.aclweb.org/anthology/P11-1008.

Bibliography XI

- Schabes, Yves und Richard C. Waters (1994). "Tree insertion grammars: a cubic-time, parsable formalism that lexicalizes context-free grammar without changing the trees produced". In: Comput. Linguist. 21, S. 479–513.
- Schützenberger, Marcel-Paul (1961). "On the definition of a family of automata". In: Information and Control 4, S. 245–270.
- Shieber, Stuart M. (2007). "Probabilistic Synchronous Tree-Adjoining Grammars for Machine Translation: The Argument from Bilingual Dictionaries". In: Proceedings of the Workshop on Syntax and Structure in Statistical Translation. Hrsg. von Dekai Wu und David Chiang.
- Shieber, Stuart M. und Yves Schabes (1990). "Synchronous Tree-Adjoining Grammars". In: Proceedings of the 13th International Conference on Computational Linguistics (COLING '90). Bd. 3, S. 253–258.

Bibliography XII

 Thatcher, James W. (1967). "Characterizing derivation trees of context-free grammars through a generalization of finite automata theory." In: J. Comput. System Sci. 1.4, S. 317–322.