Dependency-Based N-Gram Models for General Purpose Sentence Realisation

Yuqing Guo[1] Josef van Genabith[1] Haifeng Wang[2]

[1]NCLT, School of Computing
Dublin City University

[2]Research & Development Center
Toshiba (China) Co. Ltd.

COLING, Manchester
21st August, 2008

Outline

1. Introduction
   - Surface Realisation
   - Lexical-Functional Grammar
2. Dependency-Based Generation
   - Generation in LFG
   - Dependency-Based N-Gram Models
   - Generation Algorithm
3. Experiments and Results
   - Experiment Design
   - Results and Comparison

Outline

1. Introduction
   - Surface Realisation
   - Lexical-Functional Grammar
2. Dependency-Based Generation
   - Generation in LFG
   - Dependency-Based N-Gram Models
   - Generation Algorithm
3. Experiments and Results
   - Experiment Design
   - Results and Comparison

What is Surface Realisation?

Definition

A subtask of NLG to produce syntactically, morphologically, and orthographically correct sentences from a given semantic/syntactic representation.

input: lexicalised semantic or syntactic representation of linguistic content, e.g. logical forms, semantic dependency graphs, LFG f-structures

output: meaningful, grammatical and fluent natural language string

Previous Work

- Grammar-Based Generation, e.g. LFG, HPSG, CCG, TAG
  - Handcrafted rules: Penman/KPML, FUF/SURGE, Lingo/LKB, XLE etc.
    - knowledge-intensive, time-consuming and language-dependent, domain-specific
  - Wide-coverage grammars automatically extracted from treebanks: Nakanishi et al. (2005), Cahill and van Genabith (2006), White et al. (2007) etc.
    - a vast number of grammar rules due to particulars of treebanks
- Generate-and-Select Paradigm
  - N-gram language models: Nitrogen/Halogen, FERGUS, White et al. (2007) etc.
  - Log-linear feature models: Velldal and Oepen (2005), Nakanishi et al. (2005), Cahill et al. (2007) etc.
    - expensive to generate all possible alternatives

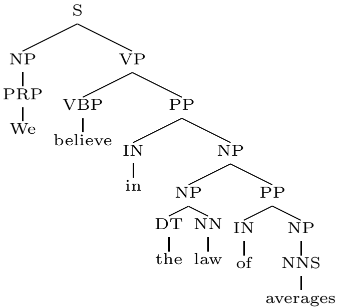
Previous Work

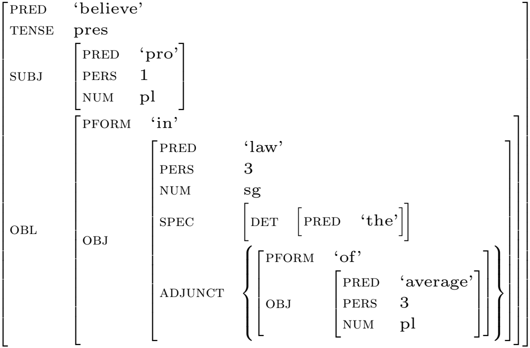
- Grammar-Based Generation, e.g. LFG, HPSG, CCG, TAG
  - Handcrafted rules: Penman/KPML, FUF/SURGE, Lingo/LKB, XLE etc.
    - knowledge-intensive, time-consuming and language-dependent, domain-specific
  - Wide-coverage grammars automatically extracted from treebanks: Nakanishi et al. (2005), Cahill and van Genabith (2006), White et al. (2007) etc.
    - a vast number of grammar rules due to particulars of treebanks
- Generate-and-Select Paradigm
  - N-gram language models: Nitrogen/Halogen, FERGUS, White et al. (2007) etc.
  - Log-linear feature models: Velldal and Oepen (2005), Nakanishi et al. (2005), Cahill et al. (2007) etc.
    - expensive to generate all possible alternatives

Alternative Solution

- Generation without Grammar
  - Maps *directly* from input representations to surface strings
  - Advantages:
    - * sidestep complex syntactic structures and a daunting number of grammar rules
    - * avoid separation between generation and selection
    - * full coverage
  - Research limited to small-scale or specialised applications
    - travel domain: Ratnaparkhi (2000), Oh and Rudnicky (2000)
    - word/phrase order: Uchimoto et al. (2000), Ringger et al. (2003), Filippova and Strube (2007)

C- and F-Structures





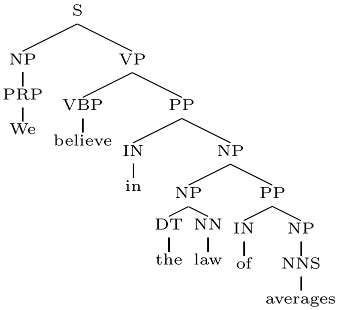
C(onstituent) Structure

phrase structure tree

F(unctional) Structure

bilexical labelled dependencies

C- and F-Structures



$$\begin{bmatrix} \text{PRED} & \text{`believe'} \\ \text{TENSE} & \text{pres} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{pl} \end{bmatrix} \\ \text{OBL} & \begin{bmatrix} \text{PFORM} & \text{`in'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`law'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{`the'} \end{bmatrix} \end{bmatrix} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PFORM} & \text{`of'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

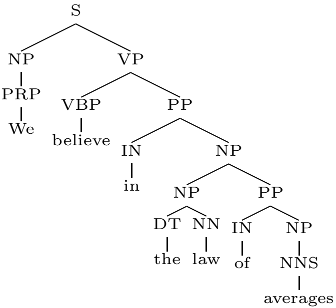
C(onstituent) Structure

phrase structure tree

F(unctional) Structure

Grammatical Functions

C- and F-Structures



$$\begin{bmatrix}
\text{PRED} & \text{`believe'} \\
\text{TENSE} & \text{pres} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{pl} \end{bmatrix} \\
\text{OBL} & \begin{bmatrix} \text{PFORM} & \text{`in'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`law'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{`the'} \end{bmatrix} \end{bmatrix} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PFORM} & \text{`of'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}
\end{bmatrix}$$

C(onstituent) Structure

phrase structure tree

F(unctional) Structure

Atomic-Valued Features

Outline

1 Introduction
- Surface Realisation
- Lexical-Functional Grammar

2 Dependency-Based Generation
- Generation in LFG
- Dependency-Based N-Gram Models
- Generation Algorithm

3 Experiments and Results
- Experiment Design
- Results and Comparison

Basic Idea

Previous Work

Determine the set of strings of the language that corresponds to a specified f-structure by applying *LFG grammar*: Cahill and van Genabith (2006), Hogan et al. (2007), Guo et al. (2008), XLE

Present Work

Determine the linear order of GFs in the given f-structure *directly* by means of n-gram models over dependency labels

A Premise for Dependency-Based Generation

- Complexity Reduction
  - selecting the most likely string in each local f-structure
- Projectivity Assumption
  - non-projective dependencies account for non-local dependencies

e.g. *How quickly did the newspapers say the athlete ran?* (Carroll et al., 2005)

$$
\left[\begin{array}{ll}
\text{FOCUS} & \left[\begin{array}{ll} \text{PRED} & \text{'quickly'} \\ \text{ADJUNCT} & \left\{\left[\begin{array}{ll} \text{PRED} & \text{'how'} \end{array}\right]\right\} \end{array}\right] \boxed{1} \\
\text{PRED} & \text{'say'} \\
\text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'newspaper'} \\ \text{SPEC} & \left[\begin{array}{ll} \text{PRED} & \text{'the'} \end{array}\right] \end{array}\right] \\
\text{COMP} & \left[\begin{array}{ll} \text{PRED} & \text{'run'} \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'athlete'} \\ \text{SPEC} & \left[\begin{array}{ll} \text{PRED} & \text{'the'} \end{array}\right] \end{array}\right] \\ & \end{array}\right]
\end{array}\right]
$$

A Premise for Dependency-Based Generation

- Complexity Reduction
  - * selecting the most likely string in each local f-structure
- Projectivity Assumption
  - * non-projective dependencies account for non-local dependencies

e.g. *How quickly did the newspapers say the athlete ran?* (Carroll et al., 2005)

$$\begin{bmatrix} \text{FOCUS} & \begin{bmatrix} \text{PRED} & \text{`quickly'} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PRED} & \text{`how'} \end{bmatrix} \right\} \end{bmatrix} \boxed{1} \\ \text{PRED} & \text{`say'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`newspaper'} \\ \text{SPEC} & \begin{bmatrix} \text{PRED} & \text{`the'} \end{bmatrix} \end{bmatrix} \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \text{`run'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`athlete'} \\ \text{SPEC} & \begin{bmatrix} \text{PRED} & \text{`the'} \end{bmatrix} \end{bmatrix} \\ \text{ADJUNCT} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

A Premise for Dependency-Based Generation

- Complexity Reduction
  - selecting the most likely string in each local f-structure
- Projectivity Assumption
  - non-projective dependencies account for non-local dependencies

e.g. *How quickly did the newspapers say the athlete ran?* (Carroll et al., 2005)

$$\left[\begin{array}{ll}
\textsc{focus} & \left[\begin{array}{ll} \textsc{pred} & \text{`quickly'} \\ \textsc{adjunct} & \left\{\left[\begin{array}{ll}\textsc{pred} & \text{`how'}\end{array}\right]\right\}\end{array}\right] \\
\textsc{pred} & \text{`say'} \\
\textsc{subj} & \left[\begin{array}{ll} \textsc{pred} & \text{`newspaper'} \\ \textsc{spec} & \left[\begin{array}{ll}\textsc{pred} & \text{`the'}\end{array}\right]\end{array}\right] \\
\textsc{comp} & \left[\begin{array}{ll} \textsc{pred} & \text{`run'} \\ \textsc{subj} & \left[\begin{array}{ll} \textsc{pred} & \text{`athlete'} \\ \textsc{spec} & \left[\begin{array}{ll}\textsc{pred} & \text{`the'}\end{array}\right]\end{array}\right]\end{array}\right]
\end{array}\right]$$

Basic N-Gram Model

$$
f_4 \left[\begin{array}{ll}
\text{PRED} & \text{‘law’} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & \left[\begin{array}{ll} \text{DET} & \left[\begin{array}{ll} \text{PRED} & \text{‘the’} \end{array} \right] \end{array} \right] \\
\text{ADJUNCT} & \left\{ \left[\begin{array}{ll} \text{PFORM} & \text{‘of’} \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{‘average’} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{array} \right] \end{array} \right] \right\}
\end{array} \right]
$$

$$\text{SPEC} \prec \text{PRED} \prec \text{ADJUNCT} \Rightarrow \text{‘the law of averages’}$$

- An (sub-)f-structure f_i containing m unordered GFs
- $S_1^m = s_1 \ldots s_m$ generated by the GF sequence $GF_1^m = GF_1 \ldots GF_m$
- N-gram model over *dependencies* searches for the best $GF_1^m = \text{argmax}\, P(GF_1^m)$

$$P(GF_1^m) = P(GF_1 ... GF_m) = \prod_{k=1}^{m} P(GF_k | GF_{k-n+1}^{k-1}) \tag{1}$$

Conditional N-Gram Models

$$
\begin{bmatrix}
\text{PFORM} & \text{`in'} \\
\text{OBJ} & f_4 \begin{bmatrix}
\text{PRED} & \text{`law'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{`the'} \end{bmatrix} \end{bmatrix} \\
\text{ADJUNCT} & \left\{ \begin{bmatrix}
\text{PFORM} & \text{`of'} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix}
\end{bmatrix} \right\}
\end{bmatrix}
\end{bmatrix}
$$

Model	N-grams			Condition
gf (P^g)	SPEC	PRED	ADJUNCT	OBJ
pred (P^p)	SPEC	PRED	ADJUNCT	'law'

$$P^g(GF_1^m) = \prod_{k=1}^{m} P(GF_k | GF_{k-n+1}^{k-1}, GF_i) \quad (2)$$

$$P^p(GF_1^m) = \prod_{k=1}^{m} P(GF_k | GF_{k-n+1}^{k-1}, Pred_i) \quad (3)$$

Conditional N-Gram Models

$$\begin{bmatrix} \text{PFORM} & \text{`in'} \\ \text{OBJ} & f_4 \begin{bmatrix} \text{PRED} & \text{`law'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{`the'} \end{bmatrix} \end{bmatrix} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PFORM} & \text{`of'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

Model	N-grams			Condition
gf (P^g)	SPEC	PRED	ADJUNCT	OBJ
pred (P^p)	SPEC	PRED	ADJUNCT	'law'

$$P^g(GF_1^m) = \prod_{k=1}^{m} P(GF_k | GF_{k-n+1}^{k-1}, GF_i) \tag{2}$$

$$P^p(GF_1^m) = \prod_{k=1}^{m} P(GF_k | GF_{k-n+1}^{k-1}, Pred_i) \tag{3}$$

Conditional N-Gram Models

$$\begin{bmatrix} \text{PFORM} & \text{`in'} \\ \text{OBJ} & f_4 \begin{bmatrix} \text{PRED} & \text{`law'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{`the'} \end{bmatrix} \end{bmatrix} \\ \text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PFORM} & \text{`of'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix} \end{bmatrix}$$

Model	N-grams			Condition
gf (P^g)	SPEC	PRED	ADJUNCT	OBJ
pred (P^p)	SPEC	PRED	ADJUNCT	'law'

$$P^g(GF_1^m) = \prod_{k=1}^{m} P(GF_k | GF_{k-n+1}^{k-1}, GF_i) \tag{2}$$

$$P^p(GF_1^m) = \prod_{k=1}^{m} P(GF_k | GF_{k-n+1}^{k-1}, Pred_i) \tag{3}$$

Featured N-Gram Models

$$
f_4 \left[\begin{array}{ll}
\text{PRED} & \text{`law'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & \left[\begin{array}{ll} \text{DET} & \left[\begin{array}{ll} \text{PRED} & \text{`the'} \end{array}\right] \end{array}\right] \\
\text{ADJUNCT} & \left\{ \left[\begin{array}{ll} \text{PFORM} & \text{`of'} \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{array}\right] \end{array}\right] \right\}
\end{array}\right]
$$

Model	N-grams		
feat (P^f)	SPEC[ ]	PRED[PERS=3,NUM=sg]	ADJUNCT[ ]
lex (P^l)	SPEC['the']	PRED['law']	ADJUNCT['of']

$$P^f(GF_1^m) = \prod_{k=1}^{m} P(Feat_k | Feat_{k-n+1}^{k-1}) \quad (4)$$

$$P^l(GF_1^m) = \prod_{k=1}^{m} P(Lex_k | Lex_{k-n+1}^{k-1}) \quad (5)$$

Featured N-Gram Models

$$
f_4 \begin{bmatrix}
\text{PRED} & \text{‘law’} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{‘the’} \end{bmatrix} \end{bmatrix} \\
\text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PFORM} & \text{‘of’} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{‘average’} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \right\}
\end{bmatrix}
$$

Model	N-grams		
feat (P^f)	SPEC[ ]	PRED[PERS=3,NUM=sg]	ADJUNCT[ ]
lex (P^l)	SPEC[‘the’]	PRED[‘law’]	ADJUNCT[‘of’]

$$P^f(GF_1^m) = \prod_{k=1}^{m} P(Feat_k | Feat_{k-n+1}^{k-1}) \tag{4}$$

$$P^l(GF_1^m) = \prod_{k=1}^{m} P(Lex_k | Lex_{k-n+1}^{k-1}) \tag{5}$$

Featured N-Gram Models

$$
f_4 \begin{bmatrix}
\text{PRED} & \text{`law'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & \begin{bmatrix} \text{DET} & \begin{bmatrix} \text{PRED} & \text{`the'} \end{bmatrix} \end{bmatrix} \\
\text{ADJUNCT} & \left\{ \begin{bmatrix} \text{PFORM} & \text{`of'} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \right\}
\end{bmatrix}
$$

Model	N-grams		
feat (P^f)	SPEC[ ]	PRED[PERS=3,NUM=sg]	ADJUNCT[ ]
lex (P^l)	SPEC['the']	PRED['law']	ADJUNCT['of']

$$P^f(GF_1^m) = \prod_{k=1}^{m} P(Feat_k | Feat_{k-n+1}^{k-1}) \tag{4}$$

$$P^l(GF_1^m) = \prod_{k=1}^{m} P(Lex_k | Lex_{k-n+1}^{k-1}) \tag{5}$$

Factored N-Gram Models

- Combined Factored Models

$$P^{CF}(GF_1^m) = \sum \lambda_i P^{F_i}(GF_1^m) \qquad (6)$$
$$where \quad \sum \lambda_i = 1$$

- Smoothing

$$\hat{P}^F(GF_1^m) = \lambda P^F(GF_1^m) + (1-\lambda)P(GF_1^m) \qquad (7)$$

Generation Algorithm

Given an input f-structure f, the core algorithm recursively traverses f in a bottom-up fashion, and at each level of sub-f-structure f_i:

1. instantiates the local predicate PRED_i of f_i;
2. calculates the linearisations of the set of GFs present at f_i by N-gram models;
3. finds the most probable GF sequence among all possible linear orders by Viterbi search;
4. generates the surface string S_i yielded by f_i according to the best GF sequence.

Generation Example

$$
f_1\begin{bmatrix}
\text{PRED} & \text{`believe'} \\
\text{TENSE} & \text{pres} \\
\text{SUBJ} & f_2\begin{bmatrix} \text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{pl} \end{bmatrix} \\
\text{OBL} & f_3\begin{bmatrix}
\text{PFORM} & \text{`in'} \\
\text{OBJ} & f_4\begin{bmatrix}
\text{PRED} & \text{`law'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & f_5\begin{bmatrix} \text{DET} & f_6\begin{bmatrix} \text{PRED} & \text{`the'} \end{bmatrix} \end{bmatrix} \\
\text{ADJUNCT} & \left\{ f_7\begin{bmatrix} \text{PFORM} & \text{`of'} \\ \text{OBJ} & f_8\begin{bmatrix} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{bmatrix} \end{bmatrix} \right\}
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

	We	believe	in	the	law	of	averages
f_8							PRED
f_7						PFORM	OBJ
f_6				PRED			
f_5				DET			
f_4				SPEC	PRED	ADJUNCT	
f_3			PFORM		OBJ		
f_2	PRED						
f_1	SUBJ	PRED	OBL				

Generation Example

$$
f_1 \left[\begin{array}{ll}
\text{PRED} & \text{`believe'} \\
\text{TENSE} & \text{pres} \\
\text{SUBJ} & f_2 \left[\begin{array}{ll} \text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{pl} \end{array} \right] \\
\text{OBL} & f_3 \left[\begin{array}{ll}
\text{PFORM} & \text{`in'} \\
\text{OBJ} & f_4 \left[\begin{array}{ll}
\text{PRED} & \text{`law'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & f_5 \left[\text{DET} \quad f_6 \left[\text{PRED} \quad \text{`the'} \right] \right] \\
\text{ADJUNCT} & \left\{ f_7 \left[\begin{array}{ll}
\text{PFORM} & \text{`of'} \\
\text{OBJ} & f_8 \left[\begin{array}{ll} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{array} \right]
\end{array} \right] \right\}
\end{array} \right]
\end{array} \right]
\end{array} \right]
$$

	We	believe	in	the	law	of	averages
f_8							PRED
f_7						PFORM	OBJ
f_6				PRED			
f_5				DET			
f_4				SPEC	PRED	ADJUNCT	
f_3			PFORM		OBJ		
f_2	PRED						
f_1	SUBJ	PRED	OBL				

Generation Example

$$
f_1\left[\begin{array}{ll}
\text{PRED} & \text{`believe'} \\
\text{TENSE} & \text{pres} \\
\text{SUBJ} & f_2\left[\begin{array}{ll} \text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{pl} \end{array}\right] \\
\text{OBL} & f_3\left[\begin{array}{ll}
\text{PFORM} & \text{`in'} \\
\text{OBJ} & f_4\left[\begin{array}{ll}
\text{PRED} & \text{`law'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & f_5\left[\begin{array}{ll} \text{DET} & f_6\left[\begin{array}{ll} \text{PRED} & \text{`the'} \end{array}\right] \end{array}\right] \\
\text{ADJUNCT} & \left\{ f_7\left[\begin{array}{ll}
\text{PFORM} & \text{`of'} \\
\text{OBJ} & f_8\left[\begin{array}{ll} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{array}\right]
\end{array}\right] \right\}
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

	We	believe	in	the	law	of	averages
f_8							PRED
f_7						PFORM	OBJ
f_6				PRED			
f_5				DET			
f_4				SPEC	PRED	ADJUNCT	
f_3			PFORM		OBJ		
f_2	PRED						
f_1	SUBJ	PRED	OBL				

Generation Example

$$
f_1\left[\begin{array}{ll}
\text{PRED} & \text{‘believe’}\\
\text{TENSE} & \text{pres}\\
\text{SUBJ} & f_2\left[\begin{array}{ll}\text{PRED} & \text{‘pro’}\\ \text{PERS} & 1\\ \text{NUM} & \text{pl}\end{array}\right]\\
\text{OBL} & f_3\left[\begin{array}{ll}
\text{PFORM} & \text{‘in’}\\
\text{OBJ} & f_4\left[\begin{array}{ll}
\text{PRED} & \text{‘law’}\\
\text{PERS} & 3\\
\text{NUM} & \text{sg}\\
\text{SPEC} & f_5\left[\begin{array}{ll}\text{DET} & f_6\left[\begin{array}{ll}\text{PRED} & \text{‘the’}\end{array}\right]\end{array}\right]\\
\text{ADJUNCT} & \left\{ f_7\left[\begin{array}{ll}
\text{PFORM} & \text{‘of’}\\
\text{OBJ} & f_8\left[\begin{array}{ll}\text{PRED} & \text{‘average’}\\ \text{PERS} & 3\\ \text{NUM} & \text{pl}\end{array}\right]
\end{array}\right] \right\}
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

	We	believe	in	the	law	of	averages
f_8							PRED
f_7						PFORM	OBJ
f_6				PRED			
f_5				DET			
f_4				SPEC	PRED	ADJUNCT	
f_3			PFORM		OBJ		
f_2	PRED						
f_1	SUBJ	PRED	OBL				

Generation Example

$$f_1\begin{bmatrix} \text{PRED} & \text{`believe'} \\ \text{TENSE} & \text{pres} \\ \text{SUBJ} & f_2\begin{bmatrix}\text{PRED} & \text{`pro'}\\ \text{PERS} & 1 \\ \text{NUM} & \text{pl}\end{bmatrix} \\ \text{OBL} & f_3\begin{bmatrix} \text{PFORM} & \text{`in'} \\ \text{OBJ} & f_4\begin{bmatrix} \text{PRED} & \text{`law'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{sg} \\ \text{SPEC} & f_5\begin{bmatrix}\text{DET} & f_6\begin{bmatrix}\text{PRED} & \text{`the'}\end{bmatrix}\end{bmatrix} \\ \text{ADJUNCT} & \left\{ f_7\begin{bmatrix} \text{PFORM} & \text{`of'} \\ \text{OBJ} & f_8\begin{bmatrix}\text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl}\end{bmatrix}\end{bmatrix} \right\} \end{bmatrix}\end{bmatrix}\end{bmatrix}$$

	We	believe	in	the	law	of	averages
f_8							PRED
f_7						PFORM	OBJ
f_6				PRED			
f_5				DET			
f_4				SPEC	PRED	ADJUNCT	
f_3			PFORM		OBJ		
f_2	PRED						
f_1	SUBJ	PRED	OBL				

Generation Example

$$
f_1\left[\begin{array}{ll}
\text{PRED} & \text{`believe'} \\
\text{TENSE} & \text{pres} \\
\text{SUBJ} & f_2\left[\begin{array}{ll} \text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{pl} \end{array}\right] \\
\text{OBL} & f_3\left[\begin{array}{ll}
\text{PFORM} & \text{`in'} \\
\text{OBJ} & f_4\left[\begin{array}{ll}
\text{PRED} & \text{`law'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & f_5\left[\begin{array}{ll} \text{DET} & f_6\left[\begin{array}{ll} \text{PRED} & \text{`the'} \end{array}\right] \end{array}\right] \\
\text{ADJUNCT} & \left\{ f_7\left[\begin{array}{ll}
\text{PFORM} & \text{`of'} \\
\text{OBJ} & f_8\left[\begin{array}{ll} \text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{array}\right]
\end{array}\right] \right\}
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

	We	believe	in	the	law	of	averages
f_8							PRED
f_7						PFORM	OBJ
f_6				PRED			
f_5				DET			
f_4				SPEC	PRED	ADJUNCT	
f_3			PFORM		OBJ		
f_2	PRED						
f_1	SUBJ	PRED	OBL				

Generation Example

$$
f_1\left[\begin{array}{ll}
\text{PRED} & \text{`believe'} \\
\text{TENSE} & \text{pres} \\
\text{SUBJ} & f_2\left[\begin{array}{ll}\text{PRED} & \text{`pro'} \\ \text{PERS} & 1 \\ \text{NUM} & \text{pl}\end{array}\right] \\
\text{OBL} & f_3\left[\begin{array}{ll}
\text{PFORM} & \text{`in'} \\
\text{OBJ} & f_4\left[\begin{array}{ll}
\text{PRED} & \text{`law'} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & f_5\left[\text{DET} \quad f_6\left[\text{PRED} \quad \text{`the'}\right]\right] \\
\text{ADJUNCT} & \left\{ f_7\left[\begin{array}{ll}
\text{PFORM} & \text{`of'} \\
\text{OBJ} & f_8\left[\begin{array}{ll}\text{PRED} & \text{`average'} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl}\end{array}\right]
\end{array}\right] \right\}
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

	We	believe	in	the	law	of	averages
f_8							PRED
f_7						PFORM	OBJ
f_6				PRED			
f_5				DET			
f_4				SPEC	PRED	ADJUNCT	
f_3			PFORM		OBJ		
f_2	PRED						
f_1	SUBJ	PRED	OBL				

Generation Example

$$
f_1\left[\begin{array}{ll}
\text{PRED} & \text{`believe'} \\
\text{TENSE} & \text{pres} \\
\text{SUBJ} & f_2\left[\begin{array}{ll} \color{red}{\text{PRED}} & \color{red}{\text{`pro'}} \\ \text{PERS} & 1 \\ \text{NUM} & \text{pl} \end{array}\right] \\
\text{OBL} & f_3\left[\begin{array}{ll}
\text{PFORM} & \color{red}{\text{`in'}} \\
\text{OBJ} & f_4\left[\begin{array}{ll}
\text{PRED} & \color{red}{\text{`law'}} \\
\text{PERS} & 3 \\
\text{NUM} & \text{sg} \\
\text{SPEC} & f_5\left[\text{DET} \quad f_6\left[\text{PRED} \quad \color{red}{\text{`the'}}\right]\right] \\
\text{ADJUNCT} & \left\{ f_7\left[\begin{array}{ll}
\text{PFORM} & \color{red}{\text{`of'}} \\
\text{OBJ} & f_8\left[\begin{array}{ll} \text{PRED} & \color{red}{\text{`average'}} \\ \text{PERS} & 3 \\ \text{NUM} & \text{pl} \end{array}\right]
\end{array}\right] \right\}
\end{array}\right]
\end{array}\right]
\end{array}\right]
$$

	We	believe	in	the	law	of	averages
f_8							PRED
f_7						PFORM	OBJ
f_6				PRED			
f_5				DET			
f_4				SPEC	PRED	ADJUNCT	
f_3			PFORM		OBJ		
f_2	PRED						
f_1	SUBJ	PRED	OBL				

Generation Example

$$
f_1\left[\begin{array}{ll}
\text{PRED} & \text{`believe'}\\
\text{TENSE} & \text{pres}\\
\text{SUBJ} & f_2\left[\begin{array}{ll}\text{PRED} & \text{`pro'}\\ \text{PERS} & 1\\ \text{NUM} & \text{pl}\end{array}\right]\\
\text{OBL} & f_3\left[\begin{array}{ll}\text{PFORM} & \text{`in'}\\ \text{OBJ} & f_4\left[\begin{array}{ll} \text{PRED} & \text{`law'}\\ \text{PERS} & 3\\ \text{NUM} & \text{sg}\\ \text{SPEC} & f_5\left[\text{DET}\quad f_6\left[\text{PRED}\quad \text{`the'}\right]\right]\\ \text{ADJUNCT} & \left\{ f_7\left[\begin{array}{ll}\text{PFORM} & \text{`of'}\\ \text{OBJ} & f_8\left[\begin{array}{ll}\text{PRED} & \text{`average'}\\ \text{PERS} & 3\\ \text{NUM} & \text{pl}\end{array}\right]\end{array}\right]\right\}\end{array}\right]\end{array}\right]
\end{array}\right]
$$

	We	believe	in	the	law	of	averages
f_8							PRED
f_7						PFORM	OBJ
f_6				PRED			
f_5				DET			
f_4				SPEC	PRED	ADJUNCT	
f_3			PFORM		OBJ		
f_2	PRED						
f_1	SUBJ	PRED	OBL				

Outline

1. Introduction
   - Surface Realisation
   - Lexical-Functional Grammar
2. Dependency-Based Generation
   - Generation in LFG
   - Dependency-Based N-Gram Models
   - Generation Algorithm
3. Experiments and Results
   - Experiment Design
   - Results and Comparison

Experimental Data

- WSJ of Penn-II Treebank (PTB)
  - Training: Sec 02-21, 39,832 sentences
  - Test: Sec 23, 2,416 sentences
  - Development: Sec 22, 1,700 sentences
- Penn Chinese Treebank (CTB6)
  - Training: 756 files, 15,663 sentences
  - Test: 84 files, 1,708 sentences
  - Development: 50 files, 1,116 sentences

Development Set	English	Chinese
num of sent	1,700	1,116
max length of sent (#words)	110	145
ave length of sent (#words)	23	31
num of local fstr	23,289	15,847
num of local fstr per sent	13.70	14.20
max length of local fstr (#gfs)	12	16
ave length of local fstr (#gfs)	**2.56**	**2.90**

Experimental Data

- WSJ of Penn-II Treebank (PTB)
  - Training: Sec 02-21, 39,832 sentences
  - Test: Sec 23, 2,416 sentences
  - Development: Sec 22, 1,700 sentences
- Penn Chinese Treebank (CTB6)
  - Training: 756 files, 15,663 sentences
  - Test: 84 files, 1,708 sentences
  - Development: 50 files, 1,116 sentences

Development Set	English	Chinese
num of sent	1,700	1,116
max length of sent (#words)	110	145
ave length of sent (#words)	23	31
num of local fstr	23,289	15,847
num of local fstr per sent	13.70	14.20
max length of local fstr (#gfs)	12	16
ave length of local fstr (#gfs)	**2.56**	**2.90**

Experimental Data

- WSJ of Penn-II Treebank (PTB)
  - Training: Sec 02-21, 39,832 sentences
  - Test: Sec 23, 2,416 sentences
  - Development: Sec 22, 1,700 sentences
- Penn Chinese Treebank (CTB6)
  - Training: 756 files, 15,663 sentences
  - Test: 84 files, 1,708 sentences
  - Development: 50 files, 1,116 sentences

Development Set	English	Chinese
num of sent	1,700	1,116
max length of sent (#words)	110	145
ave length of sent (#words)	23	31
num of local fstr	23,289	15,847
num of local fstr per sent	13.70	14.20
max length of local fstr (#gfs)	12	16
ave length of local fstr (#gfs)	**2.56**	**2.90**

Experimental Data

- WSJ of Penn-II Treebank (PTB)
  - Training: Sec 02-21, 39,832 sentences
  - Test: Sec 23, 2,416 sentences
  - Development: Sec 22, 1,700 sentences
- Penn Chinese Treebank (CTB6)
  - Training: 756 files, 15,663 sentences
  - Test: 84 files, 1,708 sentences
  - Development: 50 files, 1,116 sentences

Development Set	English	Chinese
num of sent	1,700	1,116
max length of sent (#words)	110	145
ave length of sent (#words)	23	31
num of local fstr	23,289	15,847
num of local fstr per sent	13.70	14.20
max length of local fstr (#gfs)	12	16
ave length of local fstr (#gfs)	**2.56**	**2.90**

Experiment Preparation

- F-structures automatically converted from PTB (Cahill et al., 2004) and CTB6 (Guo et al., 2007)
- Lexical macros learned for English morphological realisation

lexical macro	surface word
pred=law, num=sg, pers=3	law
pred=average, num=pl, pers=3	averages
pred=believe, num=pl, tense=pres	believe

- Dependency-based N-gram models trained by SRILM toolkit (Stolcke, 2002) with Good-Turing discounting and Katz backoff smoothing

Experiment Preparation

- F-structures automatically converted from PTB (Cahill et al., 2004) and CTB6 (Guo et al., 2007)
- Lexical macros learned for English morphological realisation

lexical macro	surface word
pred=law, num=sg, pers=3	law
pred=average, num=pl, pers=3	averages
pred=believe, num=pl, tense=pres	believe

- Dependency-based N-gram models trained by SRILM toolkit (Stolcke, 2002) with Good-Turing discounting and Katz backoff smoothing

Experiment Preparation

- F-structures automatically converted from PTB (Cahill et al., 2004) and CTB6 (Guo et al., 2007)
- Lexical macros learned for English morphological realisation

lexical macro	surface word
pred=law, num=sg, pers=3	law
pred=average, num=pl, pers=3	averages
pred=believe, num=pl, tense=pres	believe

- Dependency-based N-gram models trained by SRILM toolkit (Stolcke, 2002) with Good-Turing discounting and Katz backoff smoothing

Evaluation Metrics

Exact Match: The percentage of generated sentences that exactly match the reference.

BLEU Score: A geometric average of n-gram accuracy, adjusted by length penalty LP

$$BLEU = \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \times LP$$

Simple String Accuracy:

$$SSA = 1 - \frac{I + D + S}{R}$$

Results of English

WSJ Sec23	Without Features		
Model	ExMatch	BLEU	SSA
baseline	5.30%	0.5074	0.5729
gf	6.62%	0.5318	0.6006
pred	8.03%	0.5697	0.6073
lex	*12.87%*	*0.6741*	*0.6943*
lex+gf	12.62%	0.6611	0.6941
lex+pred	12.25%	0.6569	0.6804

WSJ Sec23	Feature Names			Feature Names & Values		
Model	ExMatch	BLEU	SSA	ExMatch	BLEU	SSA
baseline	*15.27%*	*0.6842*	*0.6948*	15.15%	0.6829	0.6915
gf	16.76%	0.6969	0.7151	16.68%	0.6977	0.7155
pred	16.72%	0.7035	0.7012	16.76%	0.7042	0.7108
lex	19.41%	0.7384	0.7476	18.96%	0.7375	0.7412
lex+gf	19.70%	0.7388	0.7498	19.74%	0.7405	0.7508
lex+pred	**19.83%**	**0.7440**	**0.7534**	19.58%	0.7422	0.7504

Table: Results for English Penn-II WSJ section 23

Results of English

WSJ Sec23	Without Features		
Model	ExMatch	BLEU	SSA
baseline	5.30%	0.5074	0.5729
gf	6.62%	0.5318	0.6006
pred	8.03%	0.5697	0.6073
lex	*12.87%*	*0.6741*	*0.6943*
lex+gf	12.62%	0.6611	0.6941
lex+pred	12.25%	0.6569	0.6804

WSJ Sec23	Feature Names			Feature Names & Values		
Model	ExMatch	BLEU	SSA	ExMatch	BLEU	SSA
baseline	*15.27%*	*0.6842*	*0.6948*	15.15%	0.6829	0.6915
gf	16.76%	0.6969	0.7151	16.68%	0.6977	0.7155
pred	16.72%	0.7035	0.7012	16.76%	0.7042	0.7108
lex	19.41%	0.7384	0.7476	18.96%	0.7375	0.7412
lex+gf	19.70%	0.7388	0.7498	19.74%	0.7405	0.7508
lex+pred	**19.83%**	**0.7440**	**0.7534**	19.58%	0.7422	0.7504

Table: Results for English Penn-II WSJ section 23

Results of English

WSJ Sec23	Without Features		
Model	ExMatch	BLEU	SSA
baseline	5.30%	0.5074	0.5729
gf	6.62%	0.5318	0.6006
pred	8.03%	0.5697	0.6073
lex	*12.87%*	*0.6741*	*0.6943*
lex+gf	12.62%	0.6611	0.6941
lex+pred	12.25%	0.6569	0.6804

WSJ Sec23	Feature Names			Feature Names & Values		
Model	ExMatch	BLEU	SSA	ExMatch	BLEU	SSA
baseline	*15.27%*	*0.6842*	*0.6948*	15.15%	0.6829	0.6915
gf	16.76%	0.6969	0.7151	16.68%	0.6977	0.7155
pred	16.72%	0.7035	0.7012	16.76%	0.7042	0.7108
lex	19.41%	0.7384	0.7476	18.96%	0.7375	0.7412
lex+gf	19.70%	0.7388	0.7498	19.74%	0.7405	0.7508
lex+pred	**19.83%**	**0.7440**	**0.7534**	19.58%	0.7422	0.7504

Table: Results for English Penn-II WSJ section 23

Results of English

WSJ Sec23	Without Features		
Model	ExMatch	BLEU	SSA
baseline	5.30%	0.5074	0.5729
gf	6.62%	0.5318	0.6006
pred	8.03%	0.5697	0.6073
lex	*12.87%*	*0.6741*	*0.6943*
lex+gf	12.62%	0.6611	0.6941
lex+pred	12.25%	0.6569	0.6804

WSJ Sec23	Feature Names			Feature Names & Values		
Model	ExMatch	BLEU	SSA	ExMatch	BLEU	SSA
baseline	*15.27%*	*0.6842*	*0.6948*	15.15%	0.6829	0.6915
gf	16.76%	0.6969	0.7151	16.68%	0.6977	0.7155
pred	16.72%	0.7035	0.7012	16.76%	0.7042	0.7108
lex	19.41%	0.7384	0.7476	18.96%	0.7375	0.7412
lex+gf	19.70%	0.7388	0.7498	19.74%	0.7405	0.7508
lex+pred	**19.83%**	**0.7440**	**0.7534**	19.58%	0.7422	0.7504

Table: Results for English Penn-II WSJ section 23

Results of Chinese

Test	Without Features		
Model	ExMatch	BLEU	SSA
baseline	8.96%	0.5752	0.5192
gf	9.54%	0.6009	0.5302
pred	10.07%	0.6180	0.5380
lex	*13.93%*	*0.6639*	*0.5961*
lex+gf	14.81%	0.6773	0.5992
lex+pred	16.04%	0.6952	0.6082

Test	Feature Names			Feature Names & Values		
Model	ExMatch	BLEU	SSA	ExMatch	BLEU	SSA
baseline	*11.77%*	*0.6160*	*0.5464*	12.30%	0.6239	0.5520
gf	12.53%	0.6391	0.5578	13.47%	0.6486	0.5660
pred	13.35%	0.6608	0.5672	14.46%	0.6720	0.5767
lex	15.16%	0.6770	0.6044	15.98%	0.6804	0.6020
lex+gf	15.52%	0.6911	0.6097	16.80%	0.6957	0.6107
lex+pred	16.22%	0.7060	0.6145	**17.51%**	**0.7123**	**0.6154**

Table: Results for Chinese CTB6 test data

Discrepancy between Languages

- Data Size: English data is more than twice of Chinese
- F-structure Representation: 50 features in English and 39 features in Chinese
- Properties of Language
  - Chinese has little inflection and case markers, resulting in a large number of coordination constructions.

(1) 投资 百万 兴建 这个 工程
invest million build this construction
'invest million yuan *to* build the construction'

(2) 改革 开放
reform opening-up
'reform and opening up'

(3) 充沛的 精力 和 敏捷的 思维
plentiful energy and quick thinking
'energetic and agile'

Comparison with Previous Work

	Coverage	ExMatch	BLEU	SSA
Langkilde (2002)	82.7%	28.2%	**0.757**	0.696
Callaway (2003)	98.7%	**49.0%**		**0.8884**
Nakanishi et al.(2005)	83.6%		0.705	
Cahill and van Genabith(2006)	98.05%		0.6651	0.6808
Hogan et al.(2007)	99.96%		0.6882	0.7092
White et al.(2007)	94.3%	6.9%	0.5768	
this paper	***100%***	***19.83%***	***0.7440***	***0.7534***

Table: Cross system comparison of results for English WSJ section 23

- Speed of our generator

  English: 0.05 sec/sent

  Chinese: 0.14 sec/sent

Comparison with Previous Work

	Coverage	ExMatch	BLEU	SSA
Langkilde (2002)	82.7%	28.2%	**0.757**	0.696
Callaway (2003)	98.7%	**49.0%**		**0.8884**
Nakanishi et al.(2005)	83.6%		0.705	
Cahill and van Genabith(2006)	98.05%		0.6651	0.6808
Hogan et al.(2007)	99.96%		0.6882	0.7092
White et al.(2007)	94.3%	6.9%	0.5768	
this paper	***100%***	***19.83%***	***0.7440***	***0.7534***

Table: Cross system comparison of results for English WSJ section 23

- Speed of our generator

  English: 0.05 sec/sent

  Chinese: 0.14 sec/sent

Summary

- N-gram models for general purpose sentence generation from labelled bilexical dependencies.
- The method is simple, accurate, broad coverage and highly efficient in practice.
- The method generalises well to different languages and data sets.

- Future work
  - Packaging equivalent sequences and generating alternative realisations
  - Exploring further combinations of conditioning context and lexicalisation
  - Applying to different languages (German, French, Arabic etc.) and dependency representations (Nivre, 2006)

Thanks & Questions!

DCU
Dublin City University
Ollscoil Chathair Bhaile Átha Cliath

National Centre for
Language Technology

TOSHIBA
Leading Innovation >>>