Exploiting a Probabilistic Hierarchical Model for Generation

Srinivas Bangalore, Owen Rambow, John Chen

{srini,rambow,jchen}@research.att.com

Columbia Statistical Generation Day

Overview

- **Natural Language Generation**
- **Related Work**
- **Our Approach: FERGUS**
- **Experiments and Results**

Typical components of generation system

- **Text Planning:**
  - **Transform communicative goal into sequence (or structure) of elementary communicative goals.**

- **Sentence Planning:**
  - **Choose linguistic resources to achieve elementary communicative goals.**

- **Realization:**
  - **Produce surface output.**

Stochastic Models in Realization

- **Our system performs (some) lexical choice, (some) syntactic choice, morphology, linearization**

- **Useful if:**
  - **Need to generate wide variety of texts (MT)**
  - **Need to quickly customize generator to new domains or genres**

Stochastic Models in Generation: Recent Approaches

- **Langkilde and Knight (INLG 1998, NAACL 2000)**

  Features:
  - **Generation from "semantic" interlingua for MT**
  - **Hand-coded rules overgenerate possible realizations for interlingua elements**
  - **Rule output composed into lattice**
  - **Linear Language Model (bigram) chooses best path through lattice**

  Problem:
  - **No guarantee of grammaticality, e.g., subject-verb agreement – but we know how to do subject-verb agreement!**

Stochastic Models in Generation: Recent Approaches (2)

- **Ratnaparkhi (NAACL-ANLP 2000)**

  Features:
  - **Generation of NPs in restricted domain (air travel) of type *flights in the evening to New York***
  - **Semantics represented as list of attribute-value pairs**
  - **Maximum entropy model trained on dependency tree and attribute sets**

  Problems:
  - **No guarantee of grammaticality**
  - **Developed and tested on very narrow domain**
  - **Subjective evaluation by only two subjects (including author)**

Stochastic Models in Generation: Recent Approaches (3)

- **Oh & Rudnicky: n-grams on language models for specific dialog acts (for $n = 5$)**
- **Stochastic Machine Translation approaches can be seen as incorporating stochastic models for generation.**

Stochastic Models in Generation: Our Approach

- **Corpus-based learning to refine syntactic and lexical choice; generate surface string**
- **Use of stochastic tree model (TM) *and* linear language model (LM)**
- **Use of pre-existing hand crafted grammar: XTAG**

Corpus Description

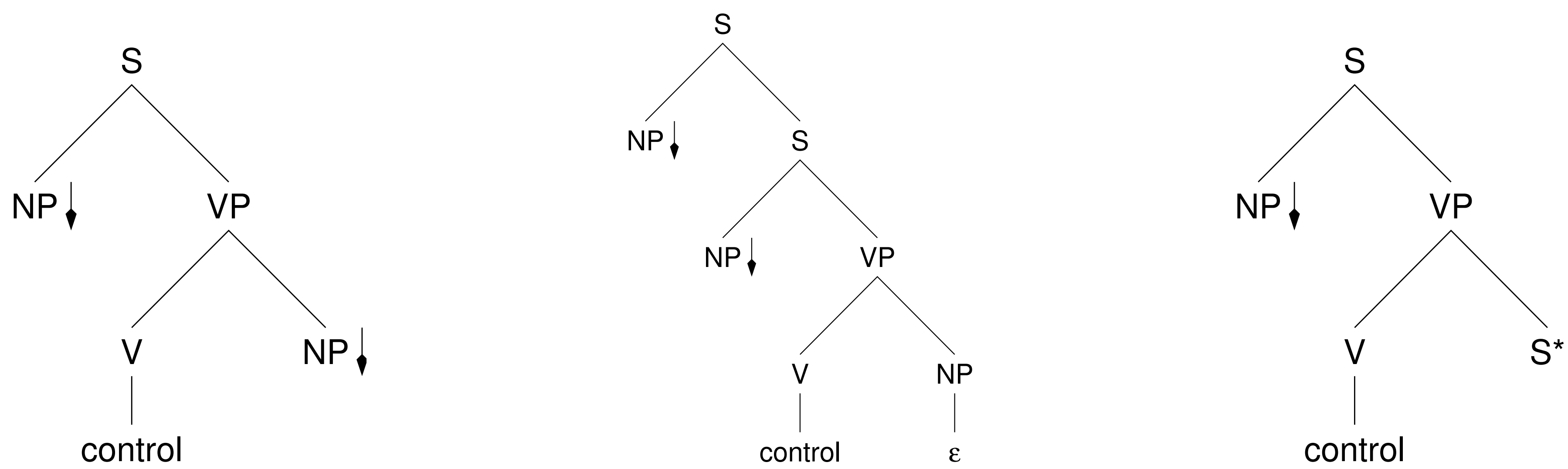
- **Dependency tree annotated with**
  - **Part-of-Speech**
  - **Morphological information**
  - **Syntactic information ("SuperTag")**
  - **Grammatical functions**
  - **Sense tags (derived from WordNet)**

Example sentence from corpus

Dep. Word	Dep POS	Dep STAG	Dep Morph	Dep LexClass	Function Words	Head Word
Treasury	NNP	A_NXN	Treasury	LC453	the	said
said	VBD	B_nx0Vs1	say_PAST_STR	LC70		-
U.S.	NNP	A_NXN	U.S.	LC256	the	default
default	VB	A_nx0V	default_INF	LC604	will	said
on	IN	B_vxPnx	on	LC150		default
Nov.	NNP	B_Nn	Nov.	LC15		ninth
ninth	CD	A_NXN	ninth	LC482		on

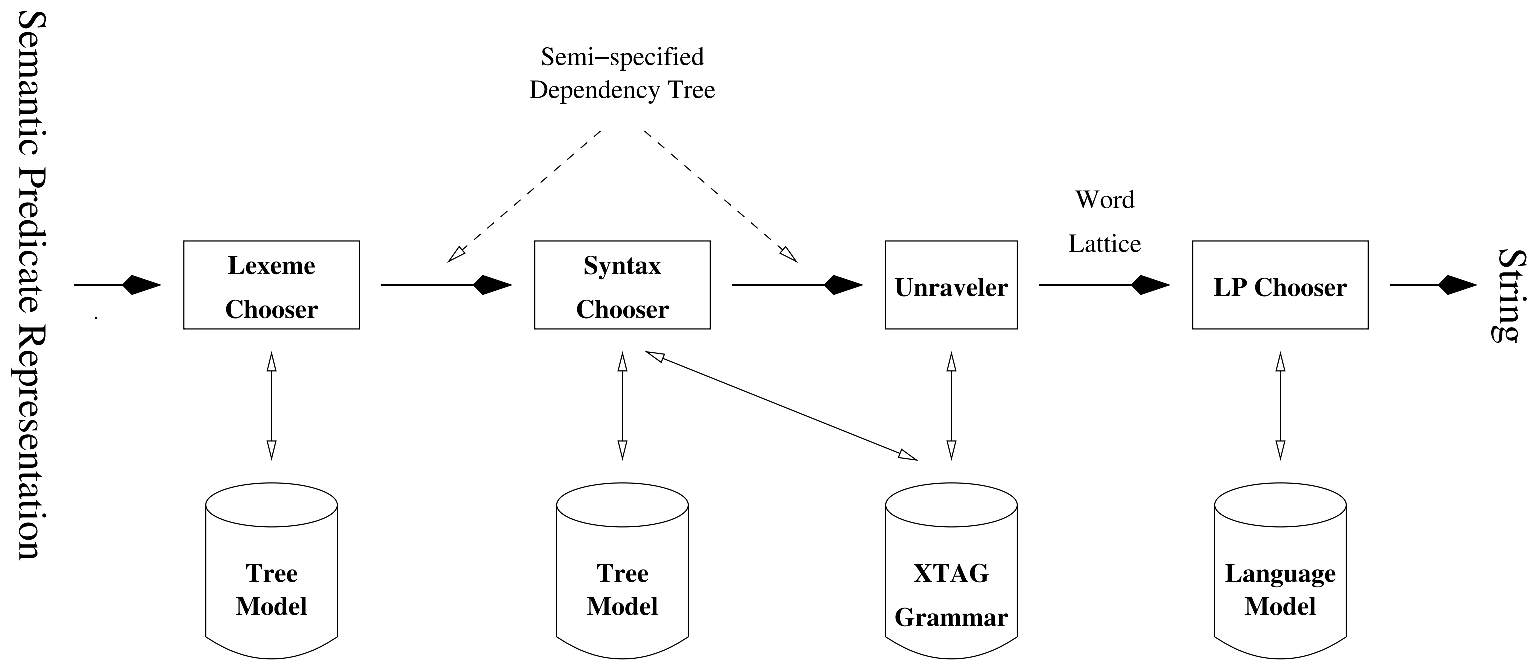
SuperTags

- **Elementary trees of Lexicalized Tree-Adjoining Grammar.**
  - **grammatical function**
  - **subcategorization frame (active valency)**
  - **passive valency with direction of modification**
  - **realizations of arguments (passive, wh-movement, relative clause)**



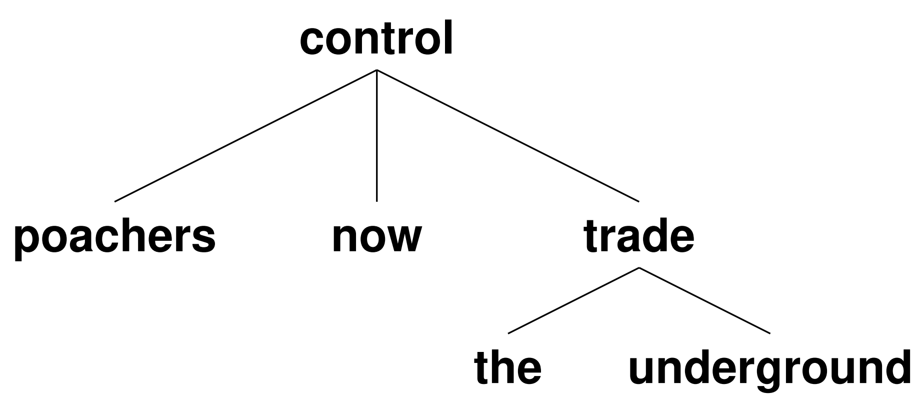
FERGUS System Architecture

- **FERGUS: Flexible Empirical/Rationalist Generation Using Syntax**



Input Representation

- **Dependency tree**
  - **any feature can be specified**

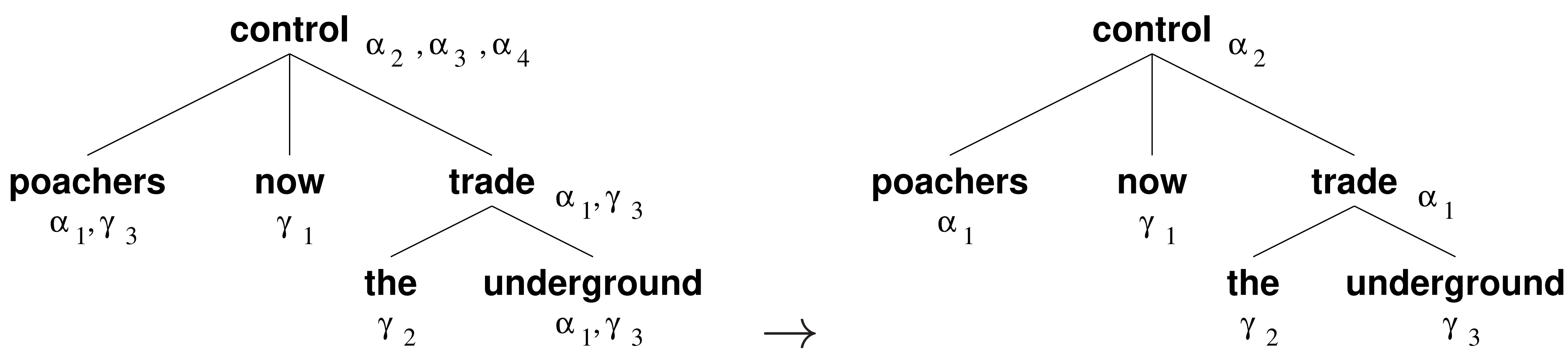


Tree Chooser

- **Given a dependency tree assign the appropriate supertag to each node**

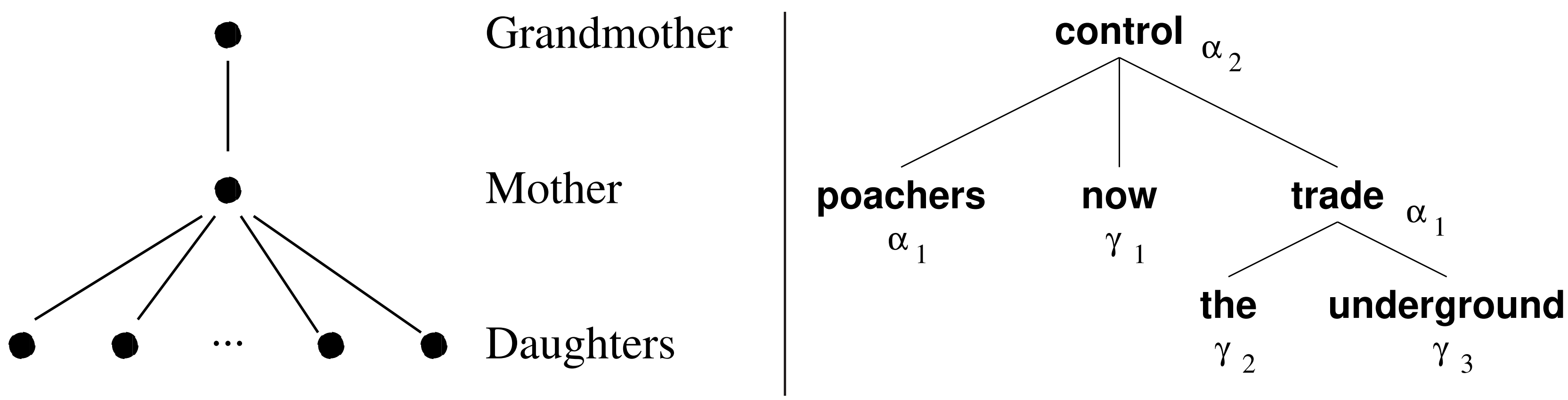
- **Probability of daughter supertag:** $P(s_d \mid s_m, l_m, l_d)$

- **Choose most probable supertag compatible with mother supertag**



Tree Chooser (2)

- **Given a dependency tree assign the appropriate supertag to each node**
- **Probability of mother supertag:** $argmax_s \, P(s \mid T_s) * P(T_s)$
- **Treelet (T_s) is modeled as a set of independent mother-daughter links.**
- **Choose most probable supertag compatible with grandmother supertag**



Treatment of Adjuncts

- **Adjunct supertags specify**
  - **category of node at which to adjoin (=passive valency)**
  - **direction of modification**

- **We collapse all adjunct (β) trees identical except for adjunction category and/or direction into one "γ-tree"**

- **Like argument trees, γ-trees only specify active valency, not passive valency**

- **Passive valency recorded in separate γ-table**

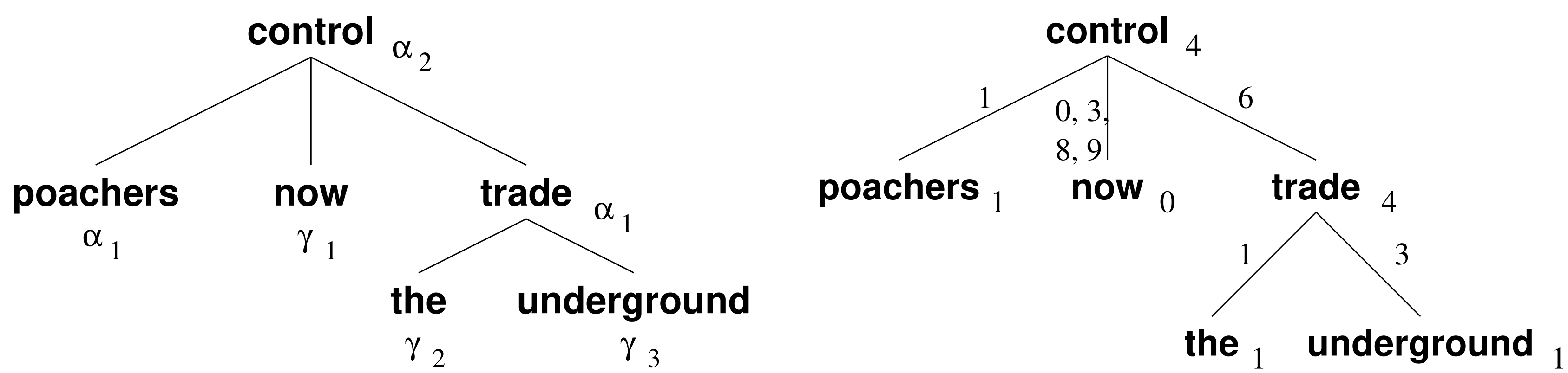
- **Similar approach in TAG chosen by McDonald (1985) and D-Tree Substitution Grammar (Rambow et.al 1995)**

Unraveler

- **Dependency tree + supertags = semi-specified derivation tree**
  - **Argument positions need not be fixed in input**
  - **Adjunction sites of γ-trees not specified**

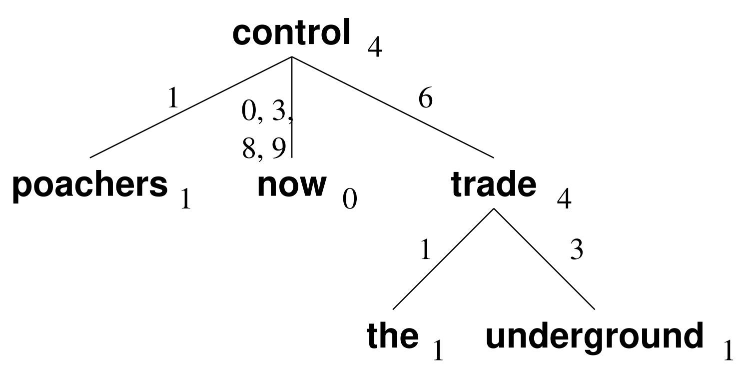
- **Given a semi-specified derivation tree, produce a word lattice**

- **XTAG grammar specifies possible positions of daughters in mother supertag.**



Unraveler (2)

- **Tree associated with supertag represented as string (= frontier)**
- **Bottom-up word lattice construction**
- **Lattice is determinized and minimized**



LP Chooser

- **Given a word lattice produce the most likely string**

- **A stochastic n-gram model is**
  - **used to rank the likelihood of the strings in the lattice**
  - **represented as a weighted finite-state automaton**
  - **composed with the word lattice**

- **Retrieves the best path through the composed lattice**

BestPath (Lattice o LangModel)

Experiment Setup

- **Baseline Model**
  - **Random dependency tree structure**
  - **Compute distribution of daughter to left (right) of mother**
- **TM-LM Model**
  - **Annotated dependency tree structure for computing LR distribution**
- **TM-XTAG Model**
  - **Annotated derivation tree structure (without gamma trees)**
  - **Compute distribution of a daughter supertag given mother information**
- **TM-XTAG-LM Model**
  - **Annotated derivation tree structure with gamma trees**

A String-Based Evaluation Metric

- **String edit distance between reference string and result string (length in words: R)**
  - **Substitutions (S)**
  - **Insertions (I)**
  - **Deletions (D)**
  - **Moves = pairs of Deletions and Insertions (M)**
  - **Remaining Insertions (I') and Deletions (D')**

	poachers	now	control	the	underground		trade	
trade	poachers			the	underground	control	now	
i	.	d	d	.	.		s	i

- **Generation String Accuracy =** $(1 - \frac{M+I'+D'+S}{R})$

Experiments and Evaluation

- **Training corpus: One million words of WSJ corpus**
- **Test corpus:**
  - **100 randomly chosen sentences**
  - **average sentence length 16.7 words**

Model	Generation String Accuracy
Baseline	0.562
TM-LM	0.668
TM-XTAG	0.684
TM-XTAG-LM (=FERGUS)	0.724

Effects of Using Automatically Created Resources

- **Grammar can be hand-crafted (XTAG) or automatically extracted from a corpus (Chen 2001)**
- **Corpus for TM can be hand-annotated (Penn Treebank) or automatically annotated (BLLiP corpus, Brown)**

Grammar	Size of Grammar	StringAccuracy
Penn Treebank (XTAG)	444	0.742
Penn Treebank (Extracted)	3063	0.749
Brown Treebank (Extracted)	3763	0.727

Effect of Using Large Corpora For Training

Can increase size of corpus for TM and LM if annotated automatically

	1M LM	2M LM
0K TM	0.685	0.703
1M TM	0.71	0.727
10M TM	0.764	0.786
20M TM	0.775	0.787
30M TM	0.767	0.784
40M TM	0.774	0.791

Conclusion

- **FERGUS uses stochastic models for surface realization**
- **Tree Model combined with Linear Model outperforms either alone**
- **Can use hand-crafted grammar or automatically extracted grammar**
- **Can use hand-annotated corpus or automatically annotated corpus (trade-off quality/quantity)**
- **Can do lexical choice during as part of syntactic choice (ACL 2000)**