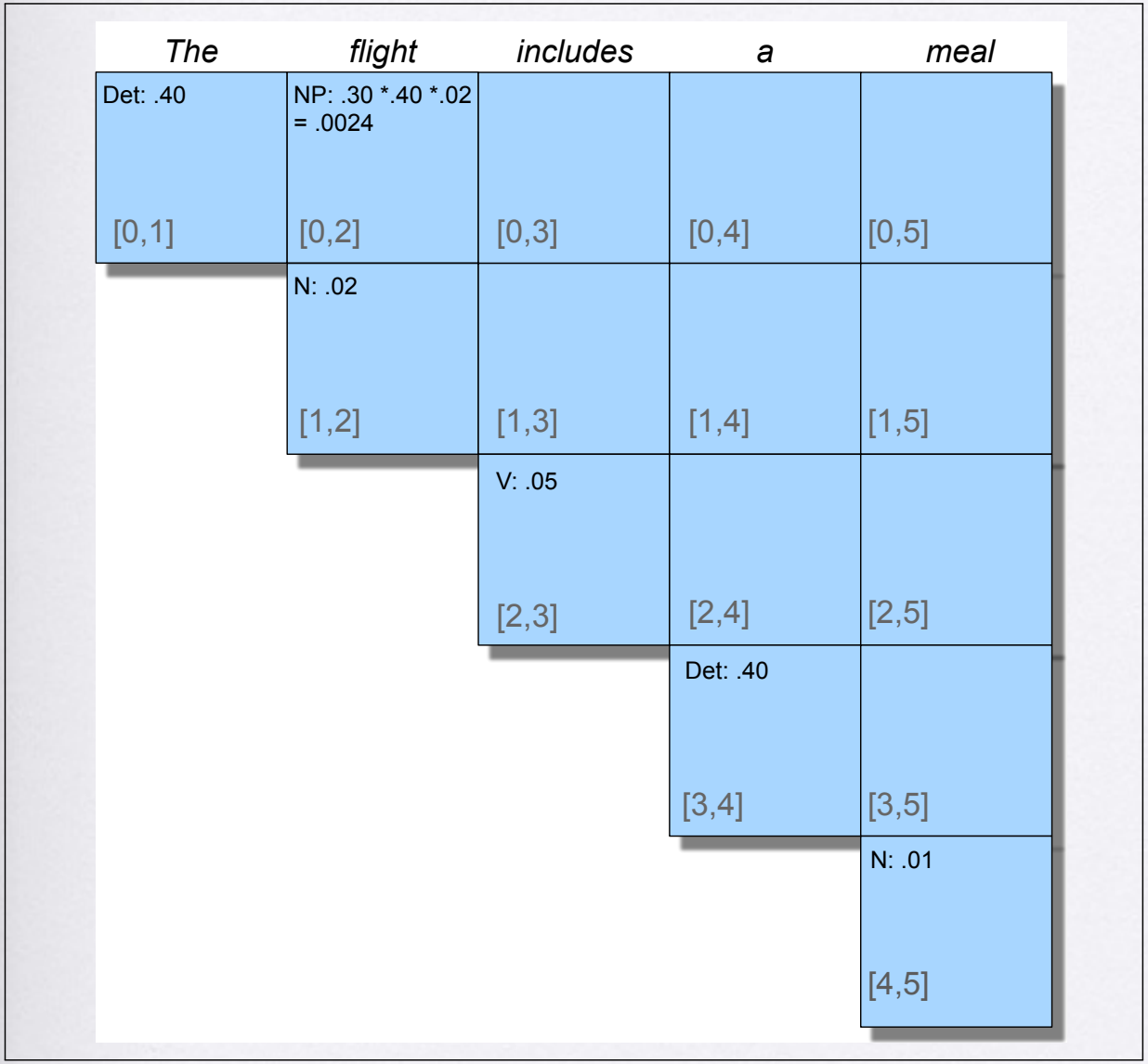
Summary of the midterm

S	$\rightarrow$	$NP\ VP$	.80	Det	$\rightarrow$	the	.40
NP	$\rightarrow$	$Det\ N$	.30	Det	$\rightarrow$	a	.40
VP	$\rightarrow$	$V\ NP$	.20	N	$\rightarrow$	$meal$	.01
V	$\rightarrow$	$includes$	.05	N	$\rightarrow$	$flight$	.02



The beginning of the probabilistic CKY matrix. Filling out the rest of the chart

G1 is a simple English grammar. The start symbol is 'S'. In order to parse sentences by CKY algorithm with this grammar, we first convert **G1** to an equivalent grammar **G2**.

1. Answer the reason why **G1** should be converted to **G2**.
2. Parse the input sentence "the boys usually build a snowman in the yard" with **G2** by CKY algorithm, and show the complete CKY table.
3. Draw all parse trees. If no parse tree is obtained, just answer 'no'.

G1

(Phrase structured rules)		(Lexical rules)	
S → NP VP	NP → D N	ADV → usually	N → snowman
S → NP ADV VP	NP → NP PP	D → a	N → yard
PP → P NP	VP → V NP	D → the	P → in
	VP → VP PP	N → boys	V → build

⇓

G2

(Phrase structured rules)		(Lexical rules)	
S → NP VP	NP → D N	ADV → usually	N → snowman
S → NP X	NP → NP PP	D → a	N → yard
X → ADV VP	VP → V NP	D → the	P → in
PP → P NP	VP → VP PP	N → boys	V → build

Answer the following questions about case analysis.

1. We consider case analysis of the following sentence S_A (where SUBJ and OBJ stands for a surface case identified by syntactic analysis).

   S_A: Prime minister (SUBJ) answered the telephone (OBJ)

   Explain the detail process of case analysis of S_A using the case frame dictionary shown in Fig.2. We suppose that the semantic primitive of 'prime minister' and 'telephone' is *human* and *concrete*, respectively. Furthermore, **give a graph** of the case structure of S_A.

	surface case	deep case	selectional restriction
answer$_1$	SUBJ	agent	human
	OBJ	patient	human/abstract
answer$_2$	SUBJ	agnet	human/org
	OBJ	patient	word
	to	goal	human
answer$_3$	SUBJ	agent	human
	OBJ	patient	concrete
answer$_4$	SUBJ	instrument	abstract
	OBJ	patient	demand

Fig. 2 Case Frame Dictionary

We consider case analysis of the sentence S_B below

S_B: Prof. Shimazu answered the phone at his room

Explain the detail process of case analysis of S_B using the example database shown in Fig.3. In the case analysis, $Sim(S_i, S_j)$, the similarity between two sentences S_i and S_j, is defined as in Equation (1):

$$Sim(S_i, S_j) \quad = \quad \text{number of common words in } S_i and S_j except for `a', `the' and `answered' \qquad (1)$$

Furthermore, **give a graph** of the case structure of S_B.

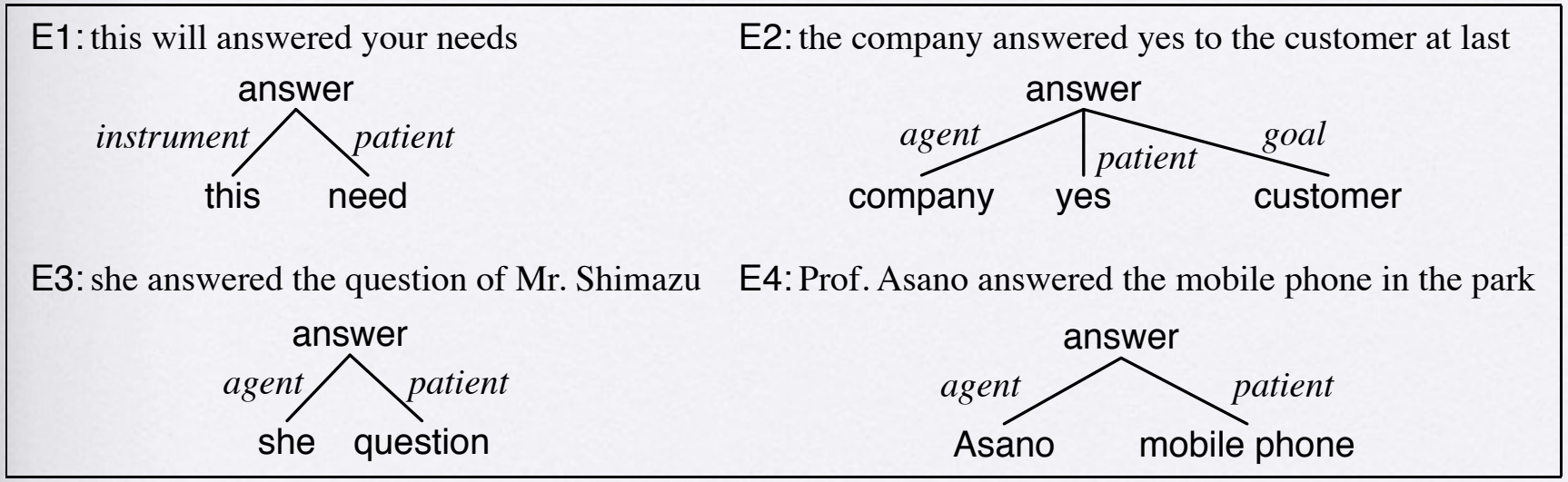


Fig. 3 Example Database

We consider POS tagging of an English sentence "go summer festival." We rank solutions of POS tagging with Hidden Markov Model(HMM), and use Viterbi algorithm to obtain the most likely solution. Fig. 3 is a word dictionary, while Fig. 4 and Fig. 5 are parameters of HMM. In the table of Fig. 5, raws stand for C_i and columns C_j. ϕ is a beginning of sentence.

Word	POS
festival	A, N
go	V
summer	A, N, V

Fig. 3 Word Dic.

$$
\begin{aligned}
P(festival|A) &= 1.5 \times 10^{-5} \\
P(festival|N) &= 5.0 \times 10^{-6} \\
P(go|V) &= 1.0 \times 10^{-5} \\
P(summer|A) &= 5.0 \times 10^{-4} \\
P(summer|N) &= 5.0 \times 10^{-5} \\
P(summer|V) &= 1.0 \times 10^{-6}
\end{aligned}
$$

Fig. 4 $P(w_i|C_i)$

	A	N	V
ϕ	0.1	0.7	0.2
A	0.1	0.8	0.1
N	0.1	0.3	0.6
V	0.3	0.6	0.1

Fig. 5 $P(C_j|C_i)$

1. Draw the word lattice. Show also the probabilities of nodes and links. Furthermore, show local best probabilities of nodes obtained by Viterbi algorithm, and links of the path corresponding to the local best probabilities.

2. Answer the most likely solution and its probability.

1. (10)

We represent the sentence S_A with semantic primitives:

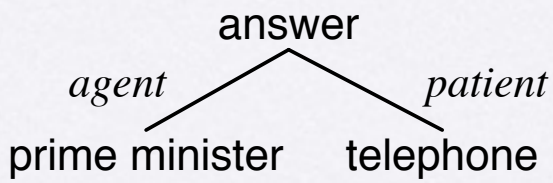
$$(\text{human})_{\text{SUBJ}} \text{ answered } (\text{location})_{\text{OBJ}}$$

answer_1: SR of case OBJ is not fulfilled.
answer_2: SRs of case OBJ and 'to' are not fulfilled.
answer_3: SR of all cases are fulfilled.
answer_4: SRs of all case are not fulfilled.
Thus we choose answer_3.
Case structure is shown below.

文 S_A を意味素で表現する

$$(\text{human})_{\text{SUBJ}} \text{ answered } (\text{location})_{\text{OBJ}}$$

answer_1: OBJ 格の選択制約が満たされない
answer_2: OBJ 格と to 格の選択制約が満たされない
answer_3: 全ての格で選択制約を満たす
answer_4: 全ての格で選択制約を満たさない
したがって、answer_3 が選ばれる。
格構造は以下の通り。



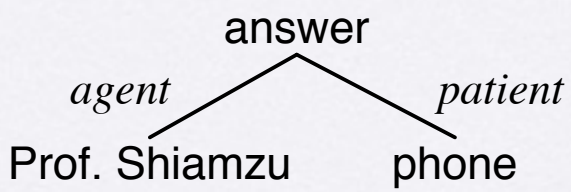
2. (10)

First, calculate similarities between S_B and each sentence in example database.
S_B と用例データベース中の文との類似度を計算する。

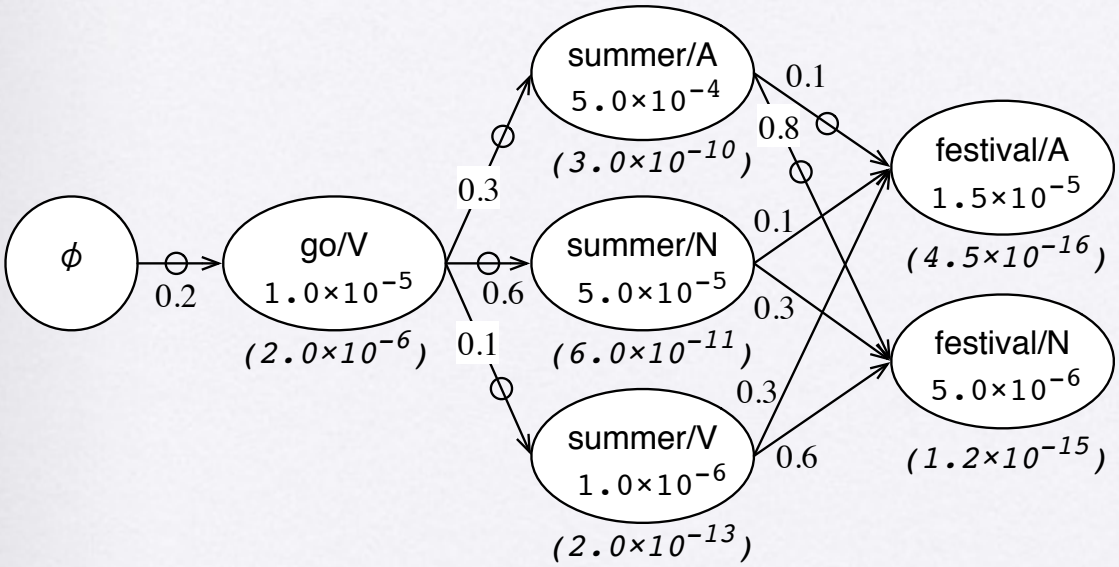
$Sim(S_B, E_1) = 0$ (no common word)
$Sim(S_B, E_2) = 1$ (common word = at)
$Sim(S_B, E_3) = 1$ (common word = Shimazu)
$Sim(S_B, E_4) = 2$ (common word = Prof., phone)

We choose E_4 whose similarity is maximum. Then we can obtain the case structure shown below by modifying the case structure of E_4.

類似度が高い E_4 を選択する。E_4 の格構造を修正して以下の格構造を得る。



1. (12)



(FYI)

- NODE festival/A

  $(3.0 \cdot 10^{-10}) \times (0.1) \times (1.5 \cdot 10^{-5}) = 4.5 \cdot 10^{-16}$ *

  $(6.0 \cdot 10^{-11}) \times (0.1) \times (1.5 \cdot 10^{-5}) = 9.0 \cdot 10^{-17}$

  $(2.0 \cdot 10^{-13}) \times (0.3) \times (1.5 \cdot 10^{-5}) = 9.0 \cdot 10^{-19}$

- NODE festival/N

  $(3.0 \cdot 10^{-10}) \times (0.8) \times (5.0 \cdot 10^{-6}) = 1.2 \cdot 10^{-15}$ *

  $(6.0 \cdot 10^{-11}) \times (0.3) \times (5.0 \cdot 10^{-6}) = 9.0 \cdot 10^{-17}$

  $(2.0 \cdot 10^{-13}) \times (0.6) \times (5.0 \cdot 10^{-6}) = 6.0 \cdot 10^{-19}$

2. (4)

go/V summer/A festival/N

1.2×10^{-15}