7ABVAGC10L$N78$j<u1Khk9=J82r@O

?9 ?.2p @>B< 2m;K 0KEl ?-BY 2.Ln ;gJf EOJU F|=PM:

F|K\ IBM E15~4pAC8&5f=j

") 242-8502 ?@F`@n8)BgOB;T2<Da4V 1623-14

1 O8aK

K\O@J8GO!"F|K\8l$rBP>]8@8l$H9k@8@.E*$J3NN(E*8@8l%b%G%k$H$=$l$K4p$E$/9=J82r@O4o$K$D$$$F=RYk!#3N%b%G%kGO!"J8$O7ABVAG$NNsH_J5$l!"$=li$N4V$G$N78$j<u$14X78$r3NN(E*J8L.<+M3J8K!$G5-=R$5lk!#3N%b%G%k$rMQ$$$?9=J82r@O4o$r:n@.$7!"F|K\7P:Q?7J9$+$iJi$k%3!<%Q%9$+$i%Q%i%a!<%?$r?dDj$7!"7ABVAGNs$rF~NOH9$k9=J82r@O<B83$r9TJC$?7k2L!"78$j<u$1C10L$G 88.6% $N2r@O@:EY$rF@$?!#$^$?!"J8;zNs$rF~NOH9$k9=J82r@O7k2L$r7ABVAG2r@O$N7k2L$H7F7FI>2A9k$H!"7ABVAG 2-gram %b%G%k$Khk7k2LhjbNI/!"9=J8>pJs$,7ABVAG2r@O$N@:EY$r8~>e$5$;$k3H$,<B83E*$K<(5l$?!#

2 78$j<u$1$K4p$E$/3NN(E*8@8l%b%G%k

3N@aGO!"2f!9$,Ds0F$9$k78$j<u1K4pE/3NN(E*8@8l%b%G%kKD$$$F=RYk!#3N%b%G%kGO!"J8$O7ABVAG$NNs ($\boldsymbol{m} = m_1 m_2 \cdots m_n$) H_J5$l!"$=li$N4VGN78$j<u$14X78$r3NN(J8L.<+M3J8K! (SCFG) $G5-=R$9$k!#=*C<5-9f$O7ABVAGG"$j!"Hs=*C<5-9f$O7ABVAG$H<u$1$k7ABVAG$N?tHNAHG"$k!#

2.1 J8$N%b%G%k

7ABVAG4V$N78$j<u1H7FCNil$k4X78$r5-=R9k$?$a$K!"0lHLE*$KG'ailF$$$kJ#?t$N78$j<u$14X78$NHs8r:9$r2>Dj$7!"7ABVAG$r=*C<5-9f$H9k3NN(J8L.<+M3J8K!$rF3F~$9$k!#F|K\8l$N78$j<u$1$N@-<A$H7F!"J8Cf$GA0$K0LCV9k7ABVAG$,!"8e$K0LCV9k7ABVAG$K78$k3H$,J,$+CF$$$k!#5i$K!"78$j<u$14X78$r!"9G$K2?$,78CF$$$k$+$K0MB8$7$J$$$Fs9`4X78H7$?2>Dj$9$H!"78$j<u$14X78$rI=$9F3=P5,B'$O $B \Rightarrow AB$ $H$$$&7A<0HJ$k!#$3$3$G!" A $O78$j7ABVAG$rI=$9!#

$9=*C<5-9f$^$?$OHs=*C<5-9fG"$j!" B $O<u$17ABVAG$rI=$9=*C<5-9f$^$?$OHs=*C<5-9f$G$"$k!#

Hs=*C<5-9f$r=*C<5-9f$HF1$8$h$&$K7ABVAGH9k3HbG-k,!"IU2CE*$J>pJs$H$ND>@Q$H9k3H$G!"%b%G%k$,78$j<u$1$N@-<A$rhjH?1G9kh&$KFC<l2=$9k3HbG-$k!#2f!9$,Ds0F9k%b%G%kGO!"$$$/DN7ABVAG$r<u$1$F$$$k$+$rIU2CE*$J>pJs$H7F2C$($k3HH7$?!#$3lO!"J8Cf$N=P8=0LCV$,6a$$7ABVAG4V$N78$j<u1O!"1s$$7ABVAG4V$N78$j<u1hjb9b$$IQEY$G@8$8$k [1] NG!"3N@-<A$r%b%G%kKAH_9~$`3?aG$"$k!#%G!<%?%9%Q!<%9%M%9$NLdBj$KBP=h9k$?aK!"<u1F$$$k7ABVAG$N?t$K>e8B$r@_$1$?!#7A<0E*Dj5A$r4JJX$K9$k$?$a$K!"=*C<5-9f$b7ABVAG$H<u$1$F$$$k7ABVAG$N?t$HNAHH9k!#?^ 1$O!"$"$kJ8$N78$j<u$19=B$$H$=$lBP1~9kJ8L.<+M3J8K!$NF3=PLZ$G$"$k!#=*C<5-9fKO78$k7ABVAG$,$J$$$H$$$&E@$KCm0U$7J1lPJi$J$$!#

0J>e$G=R$Y$?3NN(J8L.<+M3J8K!$N=*C<5-9f$N=89g T $HHs=*C<5-9f$N=89g V $O!"<u$1$F$$$k7ABVAG$N?t$r d_{max} H7$F!"0J2<$Nh&$KI=$5lk!#

$$T = \mathcal{M} \times \{0\}$$

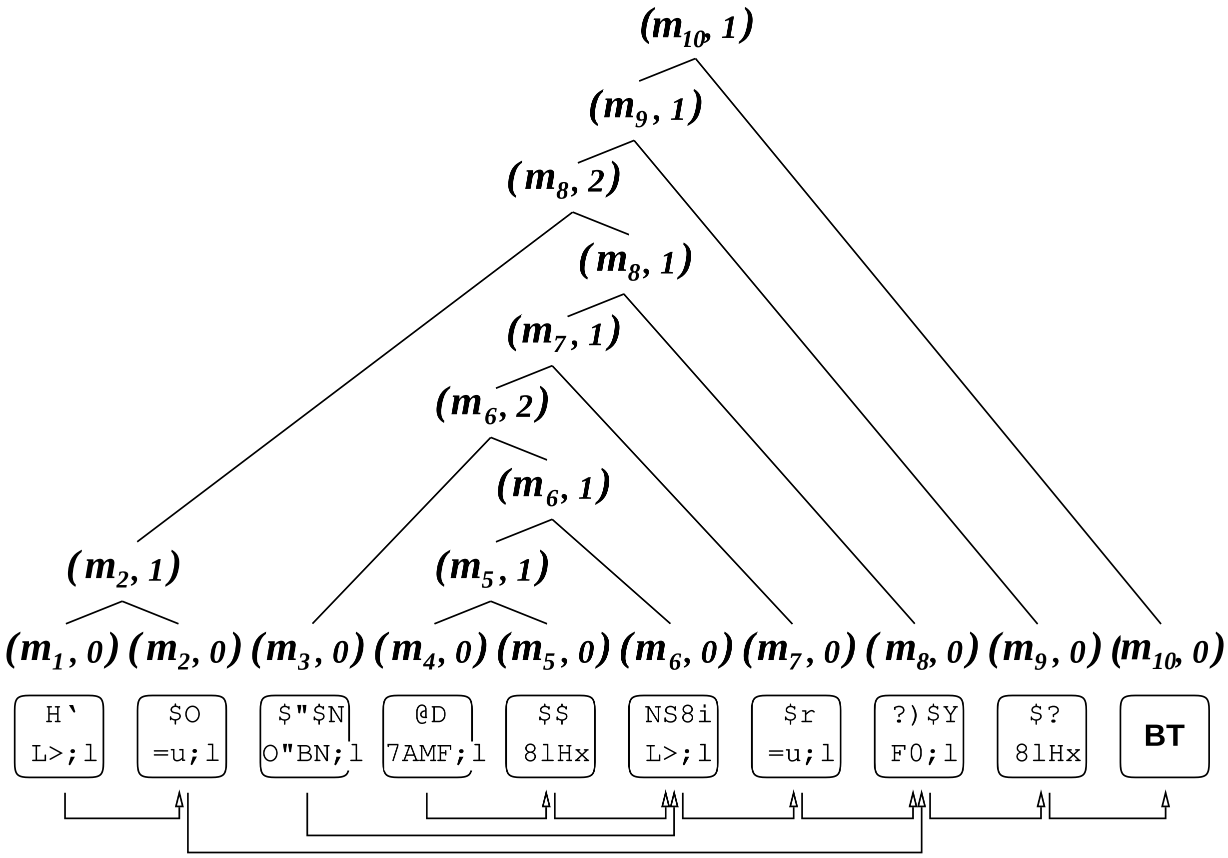
$$V = \mathcal{M} \times \{1,\ 2,\ \cdots, d_{max}\}$$

$3$3$G!" $\mathcal{M}$ $O7ABVAG$N=89g$rI=$9!#5i$K!"F3=P5,B'$O0J2<Nh$&$J7A<0KJ$k!#

$$(m_1,\ d_1) \Rightarrow (m_2,\ d_2)(m_3,\ d_3)$$

$3$3$G!"1&JU$N1&B&$N7ABVAG$,<g<-G"k=$i $m_1 = m_3$ G"$j!" 1 DN7ABVAG$r<u$1kN$G $d_1 = \min(d_3 + 1, d_{max})$ G"$k!#$^$?!"3+;O5-9f$O $(BT,\ 1)$ G"$k!#

$"$kJ8$O!"3+;O5-9f$K3liNF3=P5,B'$r2?2s$+E,MQ7?7k2LF@il$k7ABVAGNs$H7F@8@.5l$k!#3FF3=P5,B'$K$O3NN($,IUM?5lF*$j!"7ABVAGNs$N@8@.3NN(O3li$N@Q$HJk!#



?^ 1: J8L.<+M3J8K!Kh$kF3=P

2.2 DcIQEY;v>]XNBP=h

DcIQEY;v>]$KBP=h$9k?$aJd4V$rF3F~9k!#F3=P5, B′$+$iJ,$+$kh&$K!"<u$1$F$$$k7ABVAG$N?t$rL5;k9l$P!"=$>~8l$,<g<-$+$iM=B,$5lk$H$$$&0UL#G7BVAG 2-gram $KN`;w$7$F$$$k!#$7$?$,$C$F!"7ABVAG $n$-gram %b%G%k$K$*$$$F$J$5$l$k$h$&$K!"$h$jDc<!$N%b%G%k$HJd4V$9$k$3$H$GDcIQEY;v>]$KBP=h$9$k$3$H$,$G$-$k!#6qBNE*$K$O!"0J2<$N<0$,<($9$h$&$K7ABVAG 1-gram %b%G%k$H7ABVAG 0-gram %b%G%k (0lMMJ,I[) $KN`;w$9$k%b%G%k$H$NJd4V$r9TJ&!#

$$P(B \Rightarrow AB) = \lambda_2 P_2(AB|B) + \lambda_1 P_1(A) + \lambda_0 \frac{1}{|T|+|V|} \quad (1)$$

$?$@$7!" $0 \le \lambda_j \le 1$ $(j = 1, 2, 3)$ $+$D $\lambda_1 + \lambda_2 + \lambda_3 = 1$ G"$k!#$3liJd4V78?tNCMO!" n-gram %b%G%k$N>l9g$HF1$8$h$&$K!":o=|Jd4VK! [2] KhCF5aak3H$,$G$-$k!#

2.3 %Q%i%a!<%??dDj

F3=P5, B′$N3NN(CM$O!"78$j<u$1$,IUM?$5l?%3!<%Q%9$+$i$=$NIQEY$r7W?t$7!"0J2<$N<0$rMQ$$$F:GL`?dDj9k3H$GF@$ilk!#

$$P((m_1,\ d_1) \Rightarrow (m_2,\ d_2)(m_3,\ d_3)) \overset{MLE}{=} \frac{f((m_1,\ d_1) \Rightarrow (m_2,\ d_2)(m_3,\ d_3))}{f((m_1,\ d_1))}$$

$3$3$G!" $f(x)$ $O;v>] x $N3X=,%3!<%Q%9$G$NIQEY$rI=$9!#

2.4 L$CN8l%b%G%k

2f!9$N8@8l%b%G%k$,L$CN7ABVAG$r07$&$3H,G-kh$&$K!"J8;z 2-gram $+$iJkL$CN8l%b%G%k$rIU2C7?!#J8Cf$N7ABVAG$,8lWCKJ$$>l9g!"$=$NIJ;l$r=*C<$5$9f$H7F3NN(J8L.<+M3J8K!Kh$j@8@.$7!"<!$$$GL$CN8l%b%G%k$,IJ;l$+$iJ8;zNs $(x_1 x_2 \cdots x_m)$ $r0J2<

$N<0$rMQ$$$F@8@.9k!#

$$P(\boldsymbol{x}|POS) = \prod_{i=1}^{m+1} P_{POS}(x_i|x_{i-1})$$

$3$3$G!"$x_0 = x_{m+1}$ = BT $O7ABVAG$N6-3&$r<($9FC JL$N5-9f$G$"$k!#3lKh$j!"$9YF$N2DG=$JJ8;zNs KBP9$k3NN(CM$N9g7W$, 1 KJ$k!#

:o=|Jd4VKh$kJd4V78?t$N?dDjGO!"3X=,%3!<%Q%9 $rJ#?t$NItJ,%3!<%Q%9$KJ,3d$9$k!#<B83$G$O!"8lWC$O ItJ,%3!<%Q%9$N 2 $D0J>e$K=P8=9k7ABVAGH7$?!#L$ CN8l%b%G%k$N%Q%i%a!<%?$O!" 1 DNItJ,%3!<%Q%9K7 $+=P8=$7$J$$7ABVAG$+$i!"IJ;lKh$K0J2<$N<0$rMQ$$$F ?dDj5l$k!#

$$P_{\mathrm{POS}}(x_i|x_{i-1}) \stackrel{\mathrm{MLE}}{=} \frac{f_{\mathrm{POS}}(x_i, x_{i-1})}{f_{\mathrm{POS}}(x_{i-1})}$$

J8;z 2-gram %b%G%k$bJ8;z 1-gram %b%G%k$HJ8;z 0-gram %b%G%k (01MMJ,I[) $HJd4V$9$k!#Jd4V78?t$O:o=| Jd4VK! [2] Kh$j?dDj$9$k!#

3 9=J82r@O

01HLE*$K!"9=J82r@O4o$O7ABVAGNs$KJ,3d$5l?J8$r F~NO$H$7!"$=$N9=B$$r=PNO9k!#3lBP7$F!"2f!9$, Ds0F9k9=J82r@O4o$O!"J8;zNs$rF~NOH7$F!"C18l$X $NJ,3d$HIJ;l$NIUM?$r9=J82r@O$HF1;~$K9TJ&3H$, G-$k!#3N@aGO!"A0@a$G=RY?8@8l%b%G%k$K4p$E $/9=J82r@O4o$K$D$$$F@bL@9k!#

3.1 3NN(E*9=J82r@O4o

3NN(E*8@8l%b%G%k$K4p$E$/9=J82r@O4o$O!"J8;zNs $(\boldsymbol{x})$ $rM?$(ilkH!"3NN($,:GBg$HJk9=B$$r0J2<$N<0$K =>CF=PNO9k!#

$$\begin{aligned}
\hat{T} &= \operatorname*{argmax}_{\boldsymbol{w}(T)=\boldsymbol{x}} P(T|\boldsymbol{x}) \\
&= \operatorname*{argmax}_{\boldsymbol{w}(T)=\boldsymbol{x}} P(T|\boldsymbol{x})P(\boldsymbol{x}) \\
&= \operatorname*{argmax}_{\boldsymbol{w}(T)=\boldsymbol{x}} P(\boldsymbol{x}|T)P(T) \quad (\because \text{Bayes' formula}) \\
&= \operatorname*{argmax}_{\boldsymbol{w}(T)=\boldsymbol{x}} P(T) \quad (\because P(\boldsymbol{x}|T) = 1)
\end{aligned}$$

$3$3$G!" $\boldsymbol{w}(T)$ $OF3=PLZ T $N7ABVAGNs$NJ8;zNs$NO" @\$G$"$k!#:G8e$N9T$N $P(T)$ $O3NN(E*8@8l%b%G%k$G$"

I= 1: %3!<%Q%9

	J8?t	7ABVAG?t	J8;z?t
3X=,	1,072	30,292	46,212
%F%9%H	119	3,268	4,909

$k!#$3lO!"2f!9$N9=J82r@O4o$G$O!" 2@a$G=RY?3NN(J8L.<+M3J8K!KhCF7W;;5l$kF3=PLZ T $N3NN($G$" $k!#

3.2 2rC5:w$N%"%k%4%j%:%`

9=J82r@O$KMQ$$$k3NN(J8L.<+M3J8K!$NF3=P5,B'$O!" Chomsky I8=`7A$K@)8B$5$l$F$$$k!#$7$?$,$C$F!"2rC5 :w$N%"%k%4%j%:%`KOF0E*7W2hK!$N0l<o$G$"$k CKY K!$r!"3NN(J8L.<+M3J8K!$K3HD%7?%"%k%4%j%:%` [3] $r MQ$$$k3H$,$G$-$k!# CKY K!Kh$kJ8L.<+M3J8K!$N 9=J82r@O$N7W;;NL$O!"F~NO$N5-9f?t$r n H7$F $O(n^3)$ G"$k!#3NN($r07$&$?aN3HD%$O!" CKY I=$KHs=*C< 5-9fHHbK$=$3$+$iItJ,J8;zNs$,@8@.$5lk3NN($r 5-21$7F*$/$3HG<B8=5lkN$G!"7W;;NL$K$O1F6A 7J$$!#

4 I>2A

2@a$G=R$Y$?!"3NN(J8L.<+M3J8K!$K4pE/%b%G%k$r9= @.$7!" 3@a$G@bL@$7$?2rC5:w%"%k%4%j%:%`$rMQ$$$k9=J8 2r@O4o$r:n@.$7!"%F%9%H%3!<%Q%9KBP9$k9=J82r@O$N <B83$r9T$JC?!#3N@aGO!"3N7k2L$rDs<($7!"$= $NI>2A$r9TJ&!#

4.1 <B83$N>r7o

<B83KO!"F|K\7P:Q?7J9$N5-;v$+iJ$k%3!<%Q%9 (I= 1;2>H) $rMQ$$$?!#3FJ8$O!"7ABVAG$KJ,3d5l!"9=J89= B$$,IUM?5l$F$$$k!#$^$?!"%b%G%k$N@bL@$K$*$$$F$O 2DJQ$G$"$C$?!"<u$1$F$$$k7ABVAG$N?t$N>e8B$r 9 H7 $? ($d_{max} = 9$)!#

4.2 I>2A

7ABVAGC10L$N78$j<u$1%b%G%k$Khk9=J82r@O4o$r:n@.$7!"@:EY$NI>2A$r9TJC$?!#J8;zNs$KBP9k9=J82r@O$b2DG=$G$"$k$,!"7ABVAG$X$NJ,3d$N8mj,@8$8!"7k2L$NI>2A$,:$FqG"k?$a!"9=J82r@O$N@:EY$NI>2A$K$O7ABVAGNs$rF~NOH7$?>l9g$N=PNO$rMQ$$$?!#@:EY$O!"0J2<$N<0$G<($5lkh&$K!"?dDj$7$?78$j<u$14X78$N?tKBP9$k!"@5$7$$78$j<u$14X78$N3d9g$G$"$k!#$3$3$G!"@5$7$$78$j<u$14X78HO!"%3!<%Q%9$KIUM?$5l?78$j<u$14X78$HF1$8G"k3Hr0UL#9k!#

$$2r@O@:EY = \frac{78\$j@h\$,@5\$7\$\$7ABVAG\$N?t}{7ABVAG\$N?t}$$

78$j@h$,$J$$:G8e$N7ABVAG$H78$j@h$,L@i+$J:G8e$+$i 2 HVL\$N7ABVAG$O!"I>2A$NBP>]H7$F$$$J$$!#

I= 2$O!"J8L.<+M3J8K!$K4pE/%b%G%kKhkH!"9YFN7ABVAG$,<!$N7ABVAG$K78$kH9$k%Y!<%9%i%$%sHN9=J82r@O$N@:EY$G$"$k!#3X=,%3!<%Q%9$N%5%$%:$,==J,$Gk,!"%Y!<%9%i%$%s$N8mjNLsH>?t$,D{@5$5lF$$$k!#:#8e!"3X=,%3!<%Q%9$rA}$d$9$3HKhC$F@:EY8~>e$,4|BTG-$k!#$J$*!"=>Mh$NJ8@a$rC10L$H9k78$j<u$12r@O$HHf3S$K4X7F$O!"L>;l$,D>8e$N=u;l$K78$k>l9g$JINMF0W$J78$j<u1r4^$`0lJ}!"J#9g8l$N2r@O$J$I$N:$Fq$J78$j<u$1$r4^$`$H$$$&Aj0c$,$"$j!"@:EY$NHf3S$OMF0W$GOJ$$!#

$^$?!"J8;zNsKBP9$k9=J82r@O$N7k2L$r7ABVAG2r@O$N7k2LH_J7$F!"7ABVAGC10L$G$N:F8=N($HE,9gN([4]$r7ABVAG 2-gram %b%G%k$HHf3S7?!#

I= 3$O!"Ds0F%b%G%k$H7ABVAG 2-gram %b%G%k$K4p$E$/7ABVAG2r@O4o$N7k2LG"$k!#:F8=N($HE,9gN($NAPJ}$K$*$$$FDs0F%b%G%k$N@:EY$O!"7ABVAG 2-gram %b%G%kK[2]$k@:EY$hjbh/!"9=J8>pJs$,7ABVAG2r@O$N@:EY$r8~>e$5$;$k3H$,<B83E*$K<(5l$?[1]$3$1$O!"7ABVAG2r@O$H9=J82r@O$rF1;~$K9T$J$&$3HG<B8=5lF*$j!"@8@.E*$J9=J8%b%G%k$ND9=j$N 1 DG$"$k!#

5 7kO@

K\O@J8GO!"7ABVAG$rC10L$H9k78$j<u$19=B$$K4pE/3NN(E*8@8l%b%G%kKD$$$F=RY?!#7ABVAG4V$N4X78$O3NN(J8L.<+M3J8K!$G5-=R$5lk!#3N%b%G%k$rMQ$$$?9=J82r@O4o$r:n@.$7!"F|K\7P:Q?7J9$+$iJi$k 1072 J8$N%3!<%Q%9$+$i%Q%i%a!<%?$r?dDj$7!" 119 J8KBP7$F9=J82r@O<B83$r9TJC$?7k2L!"78$j<u$1C10L$?$N@:EY$O 88.6% G"C?!#$^$?!"J8;zNs$rF~NO$H9k9=J82r@O7k2L$r7ABVAG2r@O$N7k2LH7F7$FI>2A$9kH!"7ABVAG 2-gram %b%G%kKh$k7k2L$hjbh/!"9=J8>pJs$,7ABVAG2r@O$N@:EY$r8~>e$5$;$k3H$,<B83E*$K<(5l$?!#

[1]E15~9)6HBg3X$NGr0f@6><;a$O ATR BPOC%3!<%Q%9$NNcJ8 500 J8 (J?6QC18l?t 10 DxEY) KBP7$FF1MM$N7k2L$rF@$F$$$k!#

I= 2: 9=J82r@O$N@:EY

8@8l%b%G%k	2r@O@:EY
3NN(J8L.<+M3J8K!	88.6%
%Y!<%9%i%$%s*	80.0%

* $=$1$>$1$N7ABVAG$O<!$N7ABVAG$K78kH9k

I= 3: 7ABVAG2r@O$N@:EY

8@8l%b%G%k	:F8=N(	E,9gN(
3NN(J8L.<+M3J8K!	89.9% $\left(\frac{2937}{3268}\right)$	90.5% $\left(\frac{2937}{3247}\right)$
7ABVAG 2-gram	89.4% $\left(\frac{2920}{3268}\right)$	90.2% $\left(\frac{2920}{3237}\right)$

;29MJ88%

[1] Hiroshi Maruyama and Shiho Ogino. A Statistical Property of Japanese Phrase-to-Phrase Modifications. *Mathematical Linguistics*, Vol. 18, No. 7, pp. 348–352, 1992.

[2] Fredelick Jelinek, Robert L. Mercer, and Salim Roukos. Principles of Lexical Language Modeling for Speech Recognition. In *Advances in Speech Signal Processing*, chapter 21, pp. 651–699. Dekker, 1991.

[3] KL8&Fs, CfB<E/, 1JED>;L@. 2;@<8@8l=hM}. ?9KL=PHG, 1996.

[4] 1JED>;L@. EDR %3!<%Q%9$rMQ$$$?3NN(E*F|K\8l7ABVAG2r@O. EDR EE;R2=<-=qMxMQ%7%s%]%8%&%`, pp. 49–56, 1995.