# Informatics 2A: Tutorial Sheet 9 - SOLUTIONS

Mary Cryan and Shay Cohen

1. (a) Equations:

$$
\begin{aligned}
X_1 &= aX_2 + bX_1 + \epsilon \\
X_2 &= aX_1 + bX_3 \\
X_3 &= aX_2 + bX_2
\end{aligned}
$$

(b) Plugging equation for $X_3$ into that for $X_2$ we get

$$
\begin{aligned}
X_2 &= aX_1 + b(aX_2 + bX_2) \\
&= aX_1 + b(a+b)X_2 \\
&= (b(a+b))^* aX_1 && \text{by Arden's Rule}
\end{aligned}
$$

Now substituting into equation $X_1$, we get

$$
\begin{aligned}
X_1 &= a(b(a+b))^* aX_1 + bX_1 + \epsilon \\
&= (a(b(a+b))^* a + b)X_1 + \epsilon \\
&= (a(b(a+b))^* a + b)^* && \text{by Arden's Rule}
\end{aligned}
$$

Thus the regular expression for the language is

$$(a(b(a+b))^* a + b)^*$$

2. (a)

| transition | control state | unread input | stack |
|---|---|---|---|
| | $q1$ | $abba$ | $\perp$ |
| $q1 \xrightarrow{a,\perp : a\perp} q1$ | $q1$ | $bba$ | $a\perp$ |
| $q1 \xrightarrow{b,a : ba} q1$ | $q1$ | $ba$ | $b\,a\perp$ |
| $q1 \xrightarrow{\epsilon,b : b} q2$ | $q2$ | $ba$ | $b\,a\perp$ |
| $q2 \xrightarrow{b,b : \epsilon} q2$ | $q2$ | $a$ | $a\perp$ |
| $q2 \xrightarrow{a,a : \epsilon} q2$ | $q2$ | $\epsilon$ | $\perp$ |
| $q2 \xrightarrow{\epsilon,\perp : \epsilon} q2$ | $q2$ | $\epsilon$ | $\epsilon$ |

String accepted because empty stack reached at end of input.

(b) The language is:

$$\{ww' \mid w \in \{a,b\}^* \text{ and } w' \text{ is the reverse of } w\}$$

In other words, it is the language of even-length palindromes.

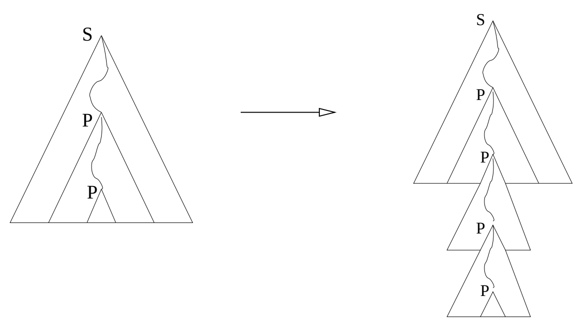
3. i. This language is regular
(only need to check that it consists of some number of $a$s, then some number of $b$s, then some number of $c$s, easily written as as a r.ex.)

ii. This language is context-free (but not regular).
(can easily adapt the pumping lemma proof for $a^m b^m$ to show this is *not* regular. Can easily adapt the CFG for $a^m b^m$ to generate this variant)

iii. This language is context-sensitive.
(can use pumping lemma for context-free languages to show it's not context sensitive)

iv. This language is context sensitive.
(this is a classic example of "not context-free", can use pumping lemma for context-free languages to show that)

The explanation behind the pumping lemma for context-free is that context-free case implies each string of the language has a parse tree. Then we observe that for any *sufficiently large* syntax tree, there will be a downward path that *visits the same non-terminal twice*. We can then 'pump in' extra copies of the relevant subtree and remain within the language:




(a)

The substring $vwx$ will be the substring generated between the two visits to the duplicate non-terminal. Then we can be sure that at least one of $v, x$ is non-empty, at least under some reasonable assumption (that we don't have circularity, say). And we can then "re-replace" the middle string $w$ by $vwx$ to create $uv^2wx^2y$, and so on ...

4. (a) The probabilistic CYK chart is as follows:

| | *fat* | *ducks* | *fish* |
|---|---|---|---|
| *fat* | N (.2), A (1.0), NP (.1) | S (.02), NP (.1) | S (.03) |
| *ducks* | | N (.4), V (.4), NP (.2), VP (.2) | S (.06), VP (.04) |
| *fish* | | | N (.4), V (.6), NP (.2), VP (.3) |

For the top right cell, there is also a competing analysis of S as (NP fat) (VP ducks fish), but this has the lower probability .004.

(b) From the construction of the above chart, the most probable parse (with probability .03) is seen to be

(S (NP (A *fat*) (NP (N *ducks*))) (VP (V *fish*))

5. (a) The semantics of *outside Scotland* (a phrase of category $\mathsf{PrP}$) is obtained as:

$$\begin{aligned} & \textit{outside}.\mathrm{Sem}(\textit{Scotland}.\mathrm{Sem}) \\ = \;& (\lambda \mathrm{y}.\lambda \mathrm{z}.\ \neg \mathrm{In}(\mathrm{z},\mathrm{y}))(\mathrm{Scotland}) \\ \to_\beta \;& (\lambda \mathrm{z}.\ \neg \mathrm{In}(\mathrm{z},\mathrm{Scotland})) \end{aligned}$$

The semantics of *city outside Scotland* (of category $\mathsf{NP}$) is therefore

$$\begin{aligned} & \lambda \mathrm{x}.\ \textit{city}.\mathrm{Sem}(\mathrm{x}) \wedge (\textit{outside Scotland}).\mathrm{Sem}(\mathrm{x}) \\ = \;& \lambda \mathrm{x}.\ ((\lambda \mathrm{z}.\mathrm{city}(\mathrm{z}))(\mathrm{x}) \wedge (\lambda \mathrm{z}.\ \neg \mathrm{In}(\mathrm{z},\mathrm{Scotland}))(\mathrm{x})) \\ \to_\beta \;& \lambda \mathrm{x}.\ \mathrm{city}(\mathrm{x}) \wedge \neg \mathrm{In}(\mathrm{x},\mathrm{Scotland}) \end{aligned}$$

Thus the semantic of the complete sentence is

$$\begin{aligned} & (\textit{city outside Scotland}).\mathrm{Sem}(\textit{London}.\mathrm{Sem}) \\ = \;& (\lambda \mathrm{x}.\ \mathrm{city}(\mathrm{x}) \wedge \neg \mathrm{In}(\mathrm{x},\mathrm{Scotland}))(\mathrm{London}) \\ \to_\beta \;& \mathrm{city}(\mathrm{London}) \wedge \neg \mathrm{In}(\mathrm{London},\mathrm{Scotland}) \end{aligned}$$

(b) The second sentence yields the logical formula

$$\neg(\mathrm{city}(\mathrm{London}) \wedge \mathrm{In}(\mathrm{London},\mathrm{Scotland}))$$

(c) The first formula implies the second by the laws of propositional logic. But not conversely: e.g. if London were not a city at all, the second formula would hold but not the first.

4. (a) Underlining the sequence to be expanded on the next line:

$$
\begin{array}{rcl}
\underline{\mathsf{S}} & \Rightarrow & \mathsf{A}_{\mathsf{hooray}}\ \underline{\ \mathsf{S}\ }\ \mathsf{E} \\
 & \Rightarrow & \mathsf{A}_{\mathsf{hooray}}\ \underline{\ \mathsf{A}_{\mathsf{hip}}\ \mathsf{E}\ }\ \mathsf{E} \\
 & \Rightarrow & \underline{\ \mathsf{A}_{\mathsf{hooray}}\ \mathsf{hip}\ }\ \mathsf{!}\ \mathsf{E} \\
 & \Rightarrow & \underline{\ \mathsf{hip}\ \mathsf{A}_{\mathsf{hooray}}\ \mathsf{!}\ }\ \mathsf{E} \\
 & \Rightarrow & \mathsf{hip}\ \mathsf{hip}\ \mathsf{!}\ \underline{\ \mathsf{A}_{\mathsf{hooray}}\ \mathsf{E}\ } \\
 & \Rightarrow & \mathsf{hip}\ \mathsf{hip}\ \mathsf{!}\ \mathsf{hooray}\ \mathsf{!}
\end{array}
$$

(b) The language $L$ consists of all phrases of the form:

$$w_1{}^n\ !\ w_2{}^{n-1}\ !\ w_3{}^{n-2}\ !\ \ldots\ w_{n-1}{}^2\ !\ w_n\ !$$

where $n \geq 1$, $w_1, \ldots, w_n$ are English words and $w^k$ means that the word $w$ is repeated $k$ times in a row, e.g., $\mathsf{hip}^2$ represents hip hip.

(c) The language is context sensitive. This follows because the grammar is an example of a noncontracting grammar.