Module Title: Informatics 2A
Exam Diet (Dec/April/Aug): Dec 2014
Brief notes on answers:

1. (a) This is the set of even-length strings, so defined by $(aa)^*$. [2]

   (b) We show the negation of the pumping property.
   Consider any $k \geq 0$.
   Choose $x = a^k b$, $y = a^k$, $z = b$. Then $xyz \in L$ and $|y| \geq k$.
   Suppose $y = uvw$ where $|v| \geq 1$.
   Choose $i = 0$.
   Then $uv^iw = uw = a^l$ for some $l < k$.
   So $xuv^iwz = a^k b a^l b \notin L$.

   L satisfies the negation of the pumping property. Hence L is not regular.

   [7 marks, in proportion to completeness/correctness]

   (c) Context sensitive [1].

2. (a) The lexing is `[, 7.10, ]` with lexical classes `[, FLT-LIT, ]` [1].

   (b) The lexer starts looking for a new lexeme with first character `7` [1]. Currently the lexical classes `INT-LIT` and `FLT-LIT` are both candidates [2].

   (c) The string `7.` is not a valid prefix of an `INT-LIT` [1]. So `7` is a completed `INT-LIT` lexeme [1]. The lexical class `FLT-LIT` is still a candidate for the current string `7.` [1].

   (d) The string `7..` is not a valid prefix of any lexical class [1]. So the lexer returns the most recent completed lexeme. This is `7` with lexical class `INT-LIT` [1], because `7.` is not in itself a valid `FLT-LIT` lexeme [1].

   [There is some flexibility about where to assign the marks above]

3. The Viterbi matrix is:

	fat	orange	ducks
N	0.5x0.2=0.1	0.12x0.6x0.3=0.0216	0.018x0.6x0.5=0.0054
V	0	0	0.0216x0.6x0.2=0.002512
A	0.3x0.4=0.12	0.12x0.3x0.5=0.018	0

   Both non-zero cells in the 'orange' column point back to (fat,A). The cell (ducks,N) points back to (orange,A), and the cell (ducks,V) to (orange,N). Thus the tagging obtained is A A N.

   [Up to 7 marks for the numbers; 2 marks for the pointers; 1 mark for the correct tagging. Minor clerical errors will not be heavily penalized if there is evidence of correct understanding.]

4. (a) LL(1) and Earley are top-down, CYK is bottom-up. [2 marks for 3 correct answers; 1 mark if 2 are correct; 0 marks otherwise.]

   (b) CYK sometimes constructs spurious parses for fragments of the sentence which are not compatible with any analysis of the sentence up to that point; Earley parsing avoids this. [1 mark]

(c) The Earley parsing table is:

S	→	• NP VP	[0,0]	P
NP	→	• N	[0,0]	P
NP	→	• the N	[0,0]	P
NP	→	things •	[0,1]	S
S	→	NP • VP	[0,1]	C
VP	→	• V	[1,1]	P
VP	→	• V N	[1,1]	P
VP	→	happen •	[1,2]	S
VP	→	happen • N	[1,2]	S
S	→	NP VP •	[0,2]	C

[Up to 7 marks. Minor variations in presentation are acceptable, e.g. writing just Nand Vin place of 'things' and 'happen', or including extra steps for N→things, V→happen.]

5. (a) The following is a suitable parameterized version of the grammar, using attribute values m,f,i for masculine, feminine, inanimate, and x as a variable ranging over these.

S	→	NP[x] VP[x]
NP[f]	→	Anna
NP[m]	→	Bill
NP[x]	→	Det N[x]
VP[x]	→	V Refl[x]
Det	→	every | some
N[f]	→	girl
N[m]	→	boy
N[i]	→	robot
V	→	hides | washes
Refl[f]	→	herself
Refl[m]	→	himself
Refl[i]	→	itself

[6 marks for the optimal solution; 4 marks for a correct but inelegant one.]

(b) The required semantic attachment is { λx. V.Sem(x,x) } [2 marks]. The expected interpretation for the given sentence is $\exists x.\ Robot(x) \wedge Washes(x, x)$ [2 marks].

6. (a) The PDA execution:

state	stack	unread input
$p1$	$\perp$	$aaab$
$p1$	a	aab
$p1$	$a\,a$	ab
$p1$	$a\,a\,a$	b
$p2$	$a\,a\,a$	b
$p2$	$a\,a$	ϵ

[6 marks: in principle 1 per step]

(b) The language is:

$$\{a^n b^m \mid 1 \leq n, 0 \leq m \leq n\}$$

[2 marks: award 1 if idea right but some error in detail]

(c) Start state $s = (p1, r1)$ [1].
Accepting states $F = \{(p2, r1)\}$ [1].
Transition relation:

$$\begin{array}{llllll} (p1,r1) & \xrightarrow{a,\bot\ :\ a} & (p1,r1) & (p1,r1) & \xrightarrow{a,a\ :\ a\,a} & (p1,r1) \\ (p1,r1) & \xrightarrow{\epsilon,a\ :\ a} & (p2,r1) & (p1,r2) & \xrightarrow{\epsilon,a\ :\ a} & (p2,r2) \\ (p2,r1) & \xrightarrow{b,a\ :\ \epsilon} & (p2,r2) & (p2,r2) & \xrightarrow{b,a\ :\ \epsilon} & (p2,r1) \end{array}$$

[6: in principle 1 per correct transition]

(d) The language is:

$$\{a^n b^m \mid 1 \leq n, 0 \leq m \leq n, \text{and } m \text{ even}\}$$

[2 marks: award 1 if idea right but some error in detail]

(e) Let M_1 be a PDA recognising L_1 and M_2 an NFA (with single start state) recognising L_2.
Let M be the product PDA as defined above.
Then M recognises $L_1 \cap L_2$.
So $L_1 \cap L_2$ is context-free, since recognised by a PDA.
[4 marks: in proportion]

(f) The languages L_1 and L_2 are both context free (as is easily shown).
Their intersection $L_1 \cap L_2$ is the language $\{a^n b^n c^n \mid n \geq 0\}$.
This language is known (and was shown in lectures) not to be context-free.
[3 marks: in proportion]

7. (a) Parse table:

	and	not	(	)	var	$
Exp		Exp1 Ops	Exp1 Ops		Exp1 Ops	
Ops	and Exp1 Ops			ϵ		ϵ
Exp1		not Exp1	(Exp)		var	

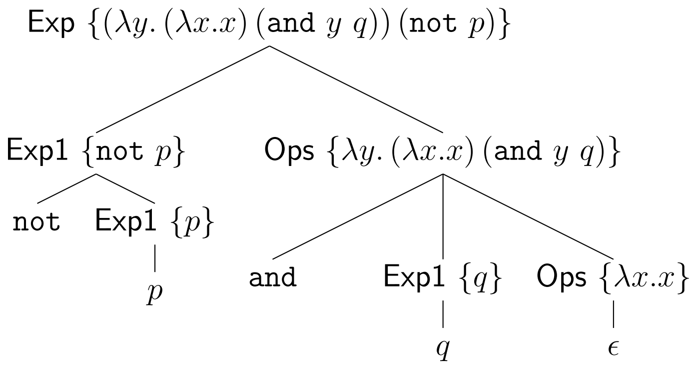
[6 marks: 1 mark penality each for up to 2 distinct kinds of mistake, otherwise mark in proportion to correctness/completeness]

(b) Algorithm execution:

action	unread input	stack
	not p and q $	Exp
Exp $\rightarrow$ Exp1 Ops	not p and q $	Exp1 Ops
Exp1 $\rightarrow$ not Exp1	not p and q $	not Exp1 Ops
match not	p and q $	Exp1 Ops
Exp1 $\rightarrow$ var	p and q $	var Ops
match var	and q $	Ops
Ops $\rightarrow$ and Exp1 Ops	and q $	and Exp1 Ops
match and	q $	Exp1 Ops
Exp1 $\rightarrow$ var	q $	var Ops
match var	$	Ops
Ops $\rightarrow$ ϵ	$	ϵ

[7 marks: in proportion]

(c) Annotated tree:



[6 marks: 2 marks each for the two annotations containing `and`; 2 for the other annotations. Penalise by just 1 mark for incorrect tree structure.]

(d) β-reductions:

$$\begin{aligned}(\lambda y.\,(\lambda x.x)\,(\texttt{and}\ y\ q))\,(\texttt{not}\ p) &\to_\beta (\lambda y.\,(\texttt{and}\ y\ q))\,(\texttt{not}\ p)\\ &\to_\beta \texttt{and not}\ p\ q\end{aligned}$$

[3 marks: 1 for having roughly right idea about β-reduction, plus 1 mark per correct step]

(e) A suitable grammar is simply:

$$\mathsf{Exp} \;\to\; \mathsf{and\ Exp\ Exp} \;\mid\; \mathsf{not\ Exp} \;\mid\; \mathsf{var}$$

[3 marks: in proportion]

8. (a) The productions are:

$$\begin{array}{rcl}
\mathsf{S} & \to & \mathsf{NP\ \ VP}\quad (1.000)\\
\mathsf{NP} & \to & \text{I}\ (3/13 = 0.231) \;\mid\; \text{me}\ (1/13 = 0.077)\\
\mathsf{NP} & \to & \mathsf{Det\ \ Nom}\quad (9/13 = 0.692)\\
\mathsf{Nom} & \to & \mathsf{Nom\ \ PP}\quad (2/11 = 0.182)\\
\mathsf{PP} & \to & \mathsf{Prep\ \ NP}\quad (1.000)\\
\mathsf{VP} & \to & \mathsf{V\ \ NP}\quad (5/6 = 0.833)\\
\mathsf{VP} & \to & \mathsf{VP\ \ PP}\quad (1/6 = 0.167)\\
\mathsf{Det} & \to & \text{a}\ (2/9 = 0.222) \;\mid\; \text{the}\ (7/9 = 0.778)\\
\mathsf{Nom} & \to & \text{dog}\ (3/11 = 0.273) \;\mid\; \text{beach}\ (1/11 = 0.091)\\
& & \mid\; \text{stick}\ (3/11 = 0.273) \;\mid\; \text{sand}\ (1/11 = 0.091)\\
& & \mid\; \text{sea}\ (1/11 = 0.091)\\
\mathsf{Prep} & \to & \text{on}\ (1/3 = 0.333) \;\mid\; \text{in}\ (1/3 = 0.333) \;\mid\; \text{towards}\ (1/3 = 0.333)\\
\mathsf{V} & \to & \text{saw}\ (3/5 = 0.600) \;\mid\; \text{threw}\ (1/5 = 0.200)\\
& & \mid\; \text{caught}\ (1/5 = 0.200)
\end{array}$$

[5 marks for the productions, 5 marks for the probabilities. Minor counting errors will not be heavily penalized where there is evidence of sound understanding.]

(b) A grammar is in CNF if the right hand side of every rule consists of either two non-terminals or a single terminal. The above grammar *is* in CNF. [2 marks]

(c) For ‘the dog saw me’, the probability is

$$1.0 \times 0.692 \times 0.778 \times 0.273 \times 0.833 \times 0.600 \times 0.077 \approx 0.00566$$

(d) The CYK chart (without explicit probabilities) is:

	I	caught	the	dog	in	the	sea
I	NP						S
caught		V		VP			VP *
the			Det	NP			NP
dog				Nom			Nom
in					Prep		PP
the						Det	NP
sea							Nom

The critical cell is the one marked *. Here we have a choice between two analyses of ‘caught the dog in the sea’ as VP :

(VP (VP caught the dog)(PP in the sea))
(VP (V caught)(NP the dog in the sea))

To see which is the more probable, note that the two parse trees involve exactly the same rules (the same number of times), except for the rule that generates the PP : in the first case VP $\rightarrow$ VP PP , and in the second case Nom $\rightarrow$ Nom PP . These rules have probabilities 0.167 and 0.182 respectively; thus the second analysis is the more probable, and this will be reflected in the pointers from the cell marked *. (In all other cases, the pointers are obvious).

[6 marks for the ordinary CYK chart. 1 mark for identifying the critical cell; 1 mark for the right choice of analysis; 2 marks for the justification.]