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 \(xuv^iw = uw = a^l\) 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 \([, \text{FLT-LIT}, ]\) [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:

<table>
<thead>
<tr>
<th></th>
<th>fat</th>
<th>orange</th>
<th>ducks</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>0.5x0.2=0.1</td>
<td>0.12x0.6x0.3=0.0216</td>
<td>0.018x0.6x0.5=0.0054</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>0</td>
<td>0.0216x0.6x0.2=0.002512</td>
</tr>
<tr>
<td>A</td>
<td>0.3x0.4=0.12</td>
<td>0.12x0.3x0.5=0.018</td>
<td>0</td>
</tr>
</tbody>
</table>

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:

\[
\begin{array}{cccc}
S & \rightarrow & \bullet \ NP \ VP & [0,0] \ P \\
NP & \rightarrow & \bullet \ N & [0,0] \ P \\
NP & \rightarrow & \bullet \ the \ N & [0,0] \ P \\
NP & \rightarrow & things \bullet & [0,1] \ S \\
S & \rightarrow & NP \bullet \ VP & [0,1] \ C \\
VP & \rightarrow & \bullet \ V & [1,1] \ P \\
VP & \rightarrow & \bullet \ V \ N & [1,1] \ P \\
VP & \rightarrow & happen \bullet & [1,2] \ S \\
VP & \rightarrow & happen \bullet \ N & [1,2] \ S \\
S & \rightarrow & NP \ VP \bullet & [0,2] \ C \\
\end{array}
\]

[Up to 7 marks. Minor variations in presentation are acceptable, e.g. writing just \texttt{N} and \texttt{V} in place of ‘things’ and ‘happen’, or including extra steps for \texttt{N} \rightarrow things, \texttt{V} \rightarrow 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.

\[
\begin{align*}
S & \rightarrow \ NP[x] \ VP[x] \\
NP[f] & \rightarrow \ Anna \\
NP[m] & \rightarrow \ Bill \\
NP[x] & \rightarrow \ Det \ N[x] \\
VP[x] & \rightarrow \ V \ Refl[x] \\
Det & \rightarrow \ every | some \\
N[f] & \rightarrow \ girl \\
N[m] & \rightarrow \ boy \\
N[i] & \rightarrow \ robot \\
V & \rightarrow \ hides | washes \\
Refl[f] & \rightarrow \ herself \\
Refl[m] & \rightarrow \ himself \\
Refl[i] & \rightarrow \ itself
\end{align*}
\]

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

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

6. (a) The PDA execution:

\[
\begin{array}{ccc}
\text{state} & \text{stack} & \text{unread input} \\
p1 & \bot & aaab \\
p1 & a & aab \\
p1 & a a & ab \\
p1 & a a a & b \\
p2 & a a a & b \\
p2 & a a & e \\
\end{array}
\]

[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 \(s = (p_1, r_1)\) [1].
Accepting states \(F = \{(p_2, r_1)\}\) [1].

Transition relation:

\[
\begin{align*}
(p_1, r_1) & \xrightarrow{a, : a} (p_1, r_1) \\
(p_1, r_1) & \xrightarrow{e, : a} (p_2, r_1) \\
(p_2, r_1) & \xrightarrow{b, : e} (p_2, r_2) \\
(p_2, r_1) & \xrightarrow{\epsilon, : a} (p_2, r_2)
\end{align*}
\]

[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:

<table>
<thead>
<tr>
<th></th>
<th>and</th>
<th>not</th>
<th>( )</th>
<th>var</th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp</td>
<td>Exp1 Ops</td>
<td>Exp1 Ops</td>
<td>Exp1 Ops</td>
<td>Exp1 Ops</td>
<td></td>
</tr>
<tr>
<td>Ops</td>
<td>and Exp1 Ops</td>
<td>$\epsilon$</td>
<td>$\epsilon$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp1</td>
<td>not Exp1</td>
<td>(Exp)</td>
<td>var</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

(b) Algorithm execution:

<table>
<thead>
<tr>
<th>action</th>
<th>unread input</th>
<th>stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp → Exp1 Ops</td>
<td>not (p) and (q) $$</td>
<td>Exp</td>
</tr>
<tr>
<td>Exp1 → not Exp1</td>
<td>not (p) and (q) $$</td>
<td>Exp1 Ops</td>
</tr>
<tr>
<td>match not</td>
<td>(p) and (q) $$</td>
<td>Exp1 Ops</td>
</tr>
<tr>
<td>Exp1 → var</td>
<td>(p) and (q) $$</td>
<td>not Exp1 Ops</td>
</tr>
<tr>
<td>match var</td>
<td>(q) $$</td>
<td>Exp1 Ops</td>
</tr>
<tr>
<td>Ops → and Exp1 Ops</td>
<td>(q) $$</td>
<td>and Exp1 Ops</td>
</tr>
<tr>
<td>match and</td>
<td>(q) $$</td>
<td>Exp1 Ops</td>
</tr>
<tr>
<td>Exp1 → var</td>
<td>(q) $$</td>
<td>var Ops</td>
</tr>
<tr>
<td>match var</td>
<td>$$</td>
<td>Ops</td>
</tr>
<tr>
<td>Ops → $\epsilon$</td>
<td>$$</td>
<td>$\epsilon$</td>
</tr>
</tbody>
</table>
[7 marks: in proportion]

(c) Annotated tree:

```
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp</td>
<td>(λy. (λx.x) (and y q)) (not p)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exp1 {not p}</td>
<td>Ops {λy. (λx.x) (and y q)}</td>
</tr>
<tr>
<td></td>
<td>not</td>
<td>Exp1 {p}</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td></td>
</tr>
<tr>
<td></td>
<td>q</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

[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:

\[
(λy. (λx.x) (\text{and } y q)) (\text{not } p) \rightarrow_β (λy. (\text{and } y q)) (\text{not } p) \\
\rightarrow_β \text{and not } p q
\]

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

(e) A suitable grammar is simply:

```
Exp \rightarrow \text{and Exp Exp} \mid \text{not Exp} \mid \text{var}
```

[3 marks: in proportion]

8. (a) The productions are:

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

[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:

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>caught</th>
<th>the</th>
<th>dog</th>
<th>in</th>
<th>the</th>
<th>sea</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>NP</td>
<td>VP</td>
<td>VP *</td>
<td>NP</td>
<td>NP</td>
<td>NP</td>
<td>S</td>
</tr>
<tr>
<td>caught</td>
<td>V</td>
<td>Det</td>
<td>Nom</td>
<td>Prep</td>
<td>Det</td>
<td>Nom</td>
<td></td>
</tr>
<tr>
<td>the</td>
<td></td>
<td>Nom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dog</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

\[
\begin{align*}
&\text{(VP (VP caught the dog)(PP in the sea))} \\
&\text{(VP (V caught)(NP the dog in the sea))}
\end{align*}
\]

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 \(\text{VP} \rightarrow \text{VP PP}\), and in the second case \(\text{Nom} \rightarrow \text{Nom PP}\). 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.]