Attempto Controlled English Language, Tools and Applications
Exercises 5

1 Translation from natural language to ACE

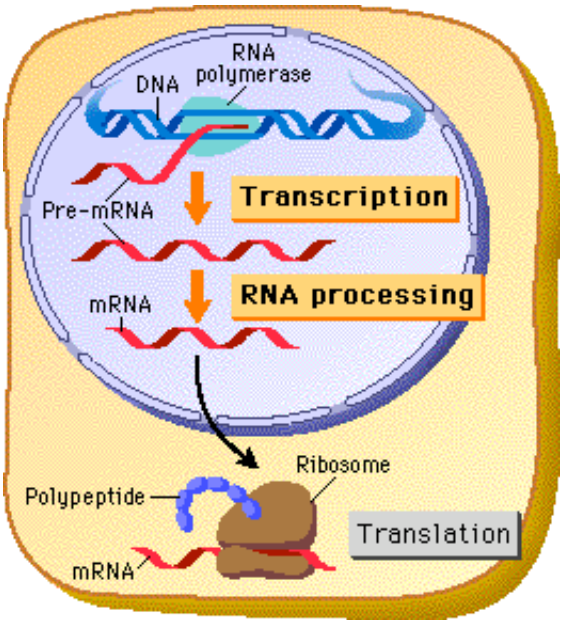
1.1 Translate from English to ACE. Real examples from molecular biology

The task in this exercise is to build a small ontology about molecular biology, capturing the basic properties of genes, proteins, etc. The original sentences are taken from Wikipedia. (Feel free to consult Wikipedia for additional domain-specific knowledge.)

Hint: Start every sentence with *if* or *every*!

Hint: Don't try to express all the sentences or all the content in the sentences.

Requirement: the resulting ontology must parse into a DRS!



1. **Deoxyribonucleic acid** (DNA) is a nucleic acid that contains the genetic instructions for the biological development of a cellular form of life or a virus. All known cellular life and some viruses have DNAs.
2. **Genes** are encoded in an organism's genome, composed of DNA or RNA, and direct the physical development and behavior of the organism.
3. Most genes contain **non-coding regions**, that do not code for the gene products.
4. **Proteins** are large organic compounds made of amino acids arranged in a linear chain and joined together between the carboxyl atom of one amino acid and the amine nitrogen of another.
5. In chemistry, an **amino acid** is any molecule that contains both amine and carboxyl functional groups.
6. An **exon** is any region of DNA within a gene that is transcribed to the final messenger RNA (mRNA) molecule.

7. **Introns** are sections of DNA that will be spliced out after transcription.
8. The **genes** of eukaryotic organisms often contain **non-coding regions** called **introns** which are removed from the **messenger RNA** in a process known as **splicing**.
9. **Transcription** is the process through which a DNA sequence is enzymatically copied by an RNA polymerase to produce a complementary RNA.
10. **RNA polymerase** (RNAP or RNApol) is an enzyme responsible for making RNA from a DNA or RNA template. RNAP accomplishes this task by constructing RNA chains through a process termed transcription.
11. In genetics, **splicing** is a modification of genetic information after transcription, in which introns are removed and exons are joined.
12. In **translation**, messenger RNA (mRNA) is decoded to produce a specific polypeptide according to the rules specified by the genetic code. This is the process that converts an mRNA sequence into a chain of amino acids that form a protein. Translation is necessarily preceded by transcription.
13. **Enzymes** are proteins that catalyze (i.e. accelerate) chemical reactions.
14. A **chromosome** is a large macromolecule into which DNA is normally packaged in a cell.
15. In the molecular sciences, a **molecule** is a sufficiently stable, electrically neutral entity composed of two or more atoms.

1.1.1 Example

There are several decisions that we have to make when translating English texts into ACE. Let's look at one example:

> Transcription is the process through which a DNA sequence is enzymatically copied by an RNA polymerase to produce a complementary RNA.

We need to decide whether we need "both directions". Or one is enough.

```
If an RNA-polymerase transcribes a DNA-sequence then the RNA-polymerase copies
the DNA-sequence and produces an RNA which complements the DNA-sequence.

If an RNA-polymerase copies a DNA-sequence and produces an RNA which complements
the DNA-sequence then the RNA-polymerase transcribes the DNA-sequence.
```

We need to decide whether some of the words are redundant in the input. E.g. 'sequence' is redundant because every DNA is a sequence. Or is a DNA-sequence a part of the full DNA? Why not say so then ...

```
If an RNA-polymerase transcribes something that is-part-of a DNA ...
```

E.g. 'enzymatically' is redundant because every polymerase is an enzyme.

```
Every RNA-polymerase is an enzyme.
If an enzyme copies something X then the enzyme copies X enzymatically.
```

All this redundancy is in terms of some background information. Does an ACE-reasoner know this background information? (Maybe it was mentioned in the previous sentence, maybe it was mentioned in an imported ontology?) If not then we should explicitly add it.

Norbert E. Fuchs, Kaarel Kaljurand {fuchs,kalju}@ifi.unizh.ch