1.   It is sometimes suggested that the meaning of a sentence is given by the set of circumstances in which that sentence would be true. How would this approach help you translate

    "Beauty is truth, truth beauty," – that is all
    Ye know on earth, and all ye need to know.

    into another language? (5 points)

Taking the suggested approach is difficult because we would have to understand what Keats meant in *Ode on a Grecian Urn*. Without understanding the first forty-eight lines it's difficult to understand what is meant by the final two lines. Even with the entire poem at our disposal it is difficult to parse. We might want to know more about Keats, and about the milieu in which he wrote the poem, his background and what influenced him. There is an almost infinite regress of questions that we could ask.

In this case does it matter that the words "Beauty is truth, truth beauty," are something communicated by an eternal urn to man? In other words, if we are to take the suggested approach and determine a set of circumstances in which the sentence is true, do we need to know that it's meant to be taken from an eternal perspective?

As I mention in the next question, translating anything of a literary nature is inherently difficult. Perhaps it would be possible to follow the suggestion with respect to the five words "Beauty is truth, truth beauty," and determine circumstances in which we could agree that those five words are true. But, I don't think that's going to help us get any closer to the meaning of the sentence, and therefore will not aid in translating this into another language.

2.   When (if ever) is translation between two natural languages impossible? When (if ever) is it possible to create a perfectly equivalent translation of a text from one language into another? Briefly explain and defend both answers. (6 points)

A perfect translation between two natural languages is impossible, if only because the transmission of meaning and information using a natural language is imperfect *without* translation, let alone *with* translation. That being said, I believe that translation between two natural languages can be thought of as a spectrum. On one end there are translations of a particular text that can be regarded as very close to perfect, and on the other end of the spectrum are translations that are completely unsatisfactory and would elicit much debate about the 'correct' translation, with no consensus.

At the end of the spectrum nearer perfection I would say are translations that are non-literary and involve concrete actions or ideas. Instruction manuals fall into this category. For other types of texts it's possible to get very good (non-controversial) translations of words or concepts that are truly shared among languages, either because the languages share a not-too-distant common ancestry, or because one language borrowed the word or concept from the other. And, once again, the more concrete the better. Translating a noun is almost always going to provide fewer opportunities for confusion than translating an adjective.

At the other end of the spectrum translating any complex idea is going to be fraught with difficulty, especially so if the idea is one that was generated in a specific cultural milieu. Without adequate knowledge of the cultural milieu how can all the connotations of the words and interpretations of the ideas possibly be communicated? Literary texts fall into this category since most literary texts have numerous references to culture, not all of which are even conscious references by the author.

3. Devise a sentence that has a clear meaning, but that you think Google Translate would struggle to translate. Explain which features of the sentence create the difficulty. (If you happen to know more than one language, you can use an example from a non-English language. In that case be sure to explain, without presupposing knowledge of that language, why Google's translation of it into English is wrong.) (3 points)

At the rugby ball all the hookers sat together.

Because Google Translate is driven by statistical machine learning it will look for words or phrases that appear more commonly together to establish which interpretation of an ambiguous word is more likely. So, it's going to have great difficulty interpreting the word 'ball' appropriately as a party and not as an object that is kicked and thrown in the game of rugby. Similarly, it will most likely have difficulty with the word 'hookers' since its usage as a position in the sport of rugby is far less common than its usage as a synonym for prostitute. However, this might not pose as much difficulty given that the word rugby appears earlier in the sentence.

4. Are there any types of translation that humans can accomplish, but that you think would be impossible for any machine translation system to cope with? Briefly explain your answer. (4 points)

I think that much of human communication is non-verbal. While it is certainly possible for machines equipped with facial recognition software to identify different facial expressions. I can imagine a 'conversation' between two individuals that is conducted solely using facial gestures and eye movements that would be impossible for a machine to translate. For instance, if the two individuals want to communicate something about a third individual in a non-verbal way (as is often done) it would be impossible for a machine to interpret the facial gestures and eye movements without knowing something about the third individual and the previously existing attitudes of the two conversational partners to that third individual.

5. Give an example of a sentence that is vague, but not ambiguous. (2 points)

"The probability that Tom studies computer science is high." Probability is a precise concept to a mathematician or statistician, but in everyday use it's a vague concept, as Daniel Kahneman points out in Thinking, Fast and Slow (page 150). Reading that sentence, we don't really know exactly how likely it is that Tom studies computer science. It is not even true that one can conclude that the actual probability is greater than 0.50. We really don't know how to interpret the word 'high'. We also know when reading this sort of sentence that we're not really talking about a mathematical probability. The word probability is being used here as a synonym for likelihood.

6. Consider the expression, "I smell coffee." Describe two different situations in which you might say this sentence: (a) a situation in which the denotation of the utterance is the most important aspect of its meaning, and (b) a situation in which some connotation of its meaning is the most important. For (b), be sure to characterize what that connotation is. Finally, (c) construct a different expression that would have the same connotation as the one you described in (b). (6 points)

A situation in which the denotation of "I smell coffee." is the most important aspect of its meaning would be one in which you can actually sense with your nose the smell of coffee, either brewing or in a coffee cup, or perhaps coffee beans just having been ground. Denotation refers to the literal or primary meaning of a word.

A situation in which the connotation of "I smell coffee." is the most important aspect of its meaning would be one in which it's the idea or feeling, in addition to the literal meaning, that is most important.

Such a situation might be one in which the speaker is referring to the association of the smell of coffee with lazy Sunday mornings, or coffee brewing on a campfire early on a crisp morning, or perhaps the idea of drinking coffee to stay up late cramming for an exam or writing a report. The connotations are numerous and not every individual will share the same connotations for coffee.

An expression that might have the same connotation as a lazy Sunday morning, or a crisp morning around a campfire, might be "I smell bacon." An expression that might have the same connotation as staying up late cramming or writing might be "burning the midnight oil".

7. Give one argument for, and one argument against the view that the meaning of the word 'dog' is the biologically defined species, *Canis familiaris*? (4 points)

An argument for the view that the meaning of the word 'dog' is the biologically defined species *Canis familiaris* is that there is a clear scientific definition of what organisms constitute the species *Canis familiaris*. If there is any doubt one can appeal to the scientific definition and examine the DNA of the organism in question. Ultimately that sort of definitive test is what we want in the definition of any word.

An argument against that view is that we use 'dog' to mean dog-like. On a safari vacation we might spot an animal and refer to it as a dog. Is it a member of *Canis familiaris?* It doesn't really matter to us or to the person we are communicating with. In other situations we might even refer to inanimate objects as dogs, such as when we give a stuffed dog to a child, and then ask "Do you like the dog?" Obviously, it's not a member of *Canis familiaris* since it's not even a living organism. But, again, this doesn't matter since there is no confusion on our part or on the part of the child we are asking.

8. Currently, there are official Wikipedias in 287 different languages. For the most part, articles in any one language are not translations of articles in any other language.

   a) Suppose your task is to build automatic translation systems to translate articles directly from any one of the official Wikipedias to the language of any other. How many distinct translation systems would you need to build? Briefly justify your answer. (2 points)

If you wanted to build 'one-step' translation systems that have no intermediary language then you would require $_{287}C_2 = 41,041$ different systems (twice that if you consider $L_1 \rightarrow L_2$ translation a distinct system from $L_2 \rightarrow L_1$ translation).

   b) Suppose, instead, you decide to invent a new, "intermediate," language for Wikipedia articles. Automatic translation of articles would then be a two-step process. First, the source article is translated from its language to the intermediate language. Second, the version in the intermediate language is translated to the intended target language. With this approach, how many distinct translation systems would you need to build? Briefly justify your answer. (2 points)

In this case, if the intermediate language were truly new and not one of the 287 languages then you would need 287 translation systems (twice that if you consider $L_1 \rightarrow$ Intermediate translation a distinct system from Intermediate $\rightarrow L_1$ translation).

   c) Suppose Wikipedia were to grow to have official versions in n languages, for some n greater than 287. Generalize your results from parts a) and b) to determine how many distinct translation systems you would need to build if you followed the approach in part a) and how many you would need to build if you followed the approach in part b). Express your results as a function of n. (3 points)

	L₁ -> L₂ translation and L₂ -> L₁ translation are a single "system"	L₁ -> L₂ translation and L₂ -> L₁ translation are two distinct "systems"
One-step with n languages	$$S(n) = \frac{n(n-1)}{2}$$	$$S(n) = n(n-1)$$
Two-step with n languages	$$S(n) = n$$	$$S(n) = 2n$$

d) In the long run, which seems the more viable approach? Briefly justify your answer. (3 points)

Clearly the option of having an intermediate language translation system seems more viable, but a better solution might be to have a few different intermediate languages (a spoke and hub model). Translating from Thai to Lao through an intermediate language that is completely different would result in a poor translation, but having an intermediate language for all East Asian languages, say, might result in much higher quality translations without resorting to quadratic growth in the number of translation systems.

e) Think of the person you know who is fluent in the most number of natural languages. There's no need to identify the person. But, how many languages does this person know? Suppose the number is three or greater. How do you think this person "translates" from one language he or she knows to another? Is it more like part a) or like part b)? Pick one of the research methods that we have discussed in this course, and briefly describe how you, as a COGS researcher, might use it to investigate the way in which this person "translates" among the languages that he or she knows. (5 points)

A friend, Ja, speaks three languages fluently and a number of others much less fluently. I have had conversations with her about language learning and her own internal mental translation experiences. She arrived in Canada from Korea as a ten-year-old. She became fluent in English, continuing to speak Korean at home. After earning a Bachelors degree in Pharmacy she decided to go to Québec to learn French and then pursue a Masters degree in Pharmacy in French. She related to me the experience of learning French and then learning pharmaceutical science in French as a constant battle of translating to English and then back to French. It was only brand new vocabulary that 'stuck' in French and required no simultaneous translation. She had never experienced that (or had forgotten that experience) with respect to English and Korean. However, there was one area where she experienced even more difficulty. For some reason she thought of numbers in Korean and did mental arithmetic in Korean. So, learning statistics in French required a translation to English and then to Korean and then all the way back again! So, in this case she was processing in Korean and using English as an intermediate language. But for the rest of her French language acquisition she was only translating to English, and for brand new vocabulary there was no translation at all, merely a direct association with the concept and the French word.

To investigate what was actually going on we might design an experiment like that described by Marian, Spivey & Hirsch where we record response times to words or cues in each of the three languages. Different parts of the experiment could have phonological distractors or semantic distractors in one or all languages. The idea would be to see if all three language centers were being activated and thus competing, or only two centers, or just one. Then we could repeat the experiment with mental arithmetic to see how the reaction times differed.

9.   Consider the following sentences: "Jack apologized to Bill, but he didn't want to." "Jack apologized
     to Bill, but he didn't hear him." What are the anaphoric pronouns in these sentences? How can a
     listener come to know what each pronoun refers to? (5 points)

In "Jack apologized to Bill, but <u>he</u> didn't want to." the underlined word 'he' is an anaphoric pronoun
referring either to Jack or to Bill. The interpretation that is most likely is that 'he' refers to Jack. The
sentence could be re-written one of two ways:

a) Jack apologized to Bill, but Jack didn't want to.
b) Jack apologized to Bill, but Bill didn't want to.

The first sentence can be understood easily, describing a reluctant Jack. The second sentence requires the
reader to imagine a situation in which Jack and Bill are meant, or have been told, to apologize to one
another. Otherwise, we don't know what Bill didn't want to do.

In "Jack apologized to Bill, but <u>he</u> didn't hear <u>him</u>." we have two anamorphic pronouns, the underlined
'he' and the underlined 'him'. The interpretation that is most likely is that 'he' refers to Bill and 'him'
refers to Jack. The sentence could be re-written in one of four ways:

a) Jack apologized to Bill, but Bill didn't hear Jack.
b) Jack apologized to Bill, but Jack didn't hear Bill.
c) Jack apologized to Bill, but Bill didn't hear Bill.
d) Jack apologized to Bill, but Jack didn't hear Jack.

The third and fourth options are easily dismissed since we don't normally think of someone having
difficulty hearing themselves, and in the case of (c) we don't even know what it is that Bill is not
supposed to have heard. The first sentence is clearly the most likely interpretation. The second sentence is
difficult to understand and once again requires the reader to construct a scenario to make some sense of it.
Perhaps, as with the first sentence we considered, Jack and Bill have been required to apologize to one
another and the writer is leaving out the fact that Bill did apologize to Jack, but Jack didn't hear Bill.
There's a lot more guesswork involved.