# F21RP – Research Methods and Project Planning – Week 2 Lab Sheet

## Literature Review

(Based on a Group Literature Exercise by Brandon Bosch from University of Nebraska)

In this exercise you will learn how to synthesise concrete paragraphs for a literature review following the principle you learned in class: “*1 paragraph = 1 idea”* starting with a strong topic sentence.

Below you will find 13 sentences taken from a section of a literature review on the topic of Natural Language Generation (NLG) that have been randomised in order. Your task is to:

1. **re-assemble** them back into a working literature review of **4-5 paragraphs**,
2. creating a **topic sentence for each paragraph**.

Remember that there is no single right way to organize these sentences and there is enough room for interpretation for each paragraph leading to a topic sentence.

### Section Title: What is Natural Language Generation?

S1: Examples include flat semantic representations, numerical data, and structured knowledge bases.

S2: In the most widely-cited survey of NLG methods to date (Reiter & Dale, 1997, 2000), NLG is characterized as ‘the sub-field of artificial intelligence and computational linguistics that is concerned with the construction of computer systems than can produce understandable texts in English or other human languages from some underlying non-linguistic representation of information’ (Reiter & Dale, 1997, p.1).

S3: For example, text summarisation was characterized above as a text-to-text application.

S4: Although the input to systems varies considerably, it is precisely the fact that such input is not – or isn’t exclusively – linguistic that is the main challenge faced by most of the systems and approaches we will consider.

S5: Clearly the former definition fits data-to-text generation better than text-to-text generation, and indeed Reiter and Dale (2000) focus exclusively on the former, helpfully, and clearly describing the rule-based approaches that dominated the field at the time.

S6: In what follows, unless otherwise specified in context, the terms ‘Natural Language Generation’ and ‘NLG’ will be used to refer to systems that generate text from non-linguistic data.

S7: More recently, generation from visual input such as image or video has become an important challenge (e.g., Mitchell et al., 2012; Kulkarni et al., 2013; Thomason et al., 2014, among many others).

S8: In contrast, text-to-text generation refer to applications such as text summarisation that take existing texts as their input, and automatically produce a new, coherent text as output.

S9: However, many approaches to text-to-text generation (especially abstractive summarisation systems, which do not extract content wholesale from the input documents) use techniques which are also used in data-to-text, as when opinions are extracted from reviews and expressed in completely new sentences (e.g., Labbe & Portet, 2012).

S10: Typically, generation of spoken utterances is closely related to dialogue management, so that management and realisation policies are sometimes learned in tandem (e.g., Rieser & Lemon, 2011).

S11: Everybody seems to agree on what the output of an NLG system should be (text), but what the exact input is can vary substantially (McDonald, 1993).

S12: For example, the generation of spoken utterances in dialogue systems (e.g., Walker et al., 2007; Rieser & Lemon, 2009; Dethlefs, 2014) is another application of NLG.

S13: Conversely, a data-to- text generation system could conceivably rely on text-to-text generation techniques for learning how to express pieces of data in different or creative ways (McIntyre & Lapata, 2009; Gatt et al., 2009; Kondadadi et al., 2013).