Appendices

We report some additional results.

A Interaction of different components

We introduced 4 components: neural planner instead of exhaustive one, adding type information, adding output verification stage, and incorporating a referring expression generation (REG). In Table 3 we report BLEU scores (Papineni et al., 2002) for all 16 combinations of components. The numbers are averages of 5 runs with different random seeds.

B REG Error Analysis

We perform further analysis of the errors of the unsupervised LM based REG module. We categorise all entities into 3 groups: (1) names of people; (2) locations (cities / counties / countries); and (3) places and objects.

For person names, the module did not produce any errors, selecting either a correct pronoun, or either the first or last name of a person, all valid refferences.

For location names, we observe two distinct error types, both relating to our module's restriction to predict a single MASK token. The first type is in cases like "city, country" or "county, country", where the more specific location is not in the LM vocabulary, and cannot be predicted with a single token. For example, in "Punjab, Pakistan", Punjab is not contained in the vocabulary as a single token, causing the model to select "Pakistan", which we consider a mistake. The second type is when a city name is longer than a single token, as in "New York". While it is common to refer to "New Jersey" as "Jersey", it is wrong to refer to "New York" as either "New" or "York", and as BERT can only fill in one MASK token, it chooses only one (in this case "York").

Finally, for places and objects, we also identify to mistake types. The first occurs for multi-token entities. While for some cases it is possible to select the correct one (i.e., "Agra Airport" → "The Airport" or "Boston University" → "The University"), in other cases it is not possible (i.e., "Baked Alaska", where choosing either word does not produce a useful reference). The second type occurs with names of objects, like books titles. For example, for the entity "A Severed Wasp" we would like the model to predict "The Book". However, as we only allow either pronouns or words from the original entity, the model cannot produce "The book", producing the erroneous "The Wasp" instead.

C Output Examples

The following output examples demonstrate the kinds of texts produces by the final system. The following outputs are correct, expressing all and only the facts from their input graphs. We enumerate them as number of facts:

1. The leader of **Azerbaijan** is **Artur Rasizade**.
2. **Baked Alaska**, containing **Sponge Cake**, is from **France**.
3. **Above The Veil**, written by **Garth Nix**, is available in **Hardcover** and has **248** pages.
4. The **Akita Museum Of Art** is located in **Japan** where the **Brazilians In Japan** are an ethnic group. **The Museum** is located in **Akita, Akita** which is part of **Akita Prefecture** .
5. The **AWH Engineering College** in **Kuttikkattoor**, **Kerala** has **Mah, India** to its northwest . **The College** was established in **2001** and has a staff of **250**.

An example where the system failed, producing a wrong lexicalization of a fact is: "The **AWH Engineering College** is located in the state of **Kerala**, **Kochi**, in **India**. The largest city in **India** is

		Exhaustive Planning		Neural Planning	
		-	REG	-	REG
No types	No Verification	46.882	47.338	46.506	47.124
	Verified Output	46.896	47.392	46.412	47.05
With Typing	No Verification	46.194	46.768	45.902	46.628
	Verified Output	46.072	46.614	46.166	46.834

Table 3: Average BLEU score for every combination of methods (avg of 5 independent runs).

Mumbai and the river is the **Ganges**". In this example, the input entity **Kochi** refers to the leader of **Kerala**, and not tpo the location (although there is also a location by that name). The text lexicalizes this fact such that **Kerala** and **Kochi** are related, but with a relation of *part-of*, implying **Kerala** is in **Kochi**.