We propose “attribute phrases” — short descriptions describing the differences between two images (or two categories).

**Attributes**:
- Example: Is a military plane? Yes (A) No (B)
- Modular — an instance can be described by a set of attributes
- Fixed set of attributes designed by experts

**Listener Model**:
- A set of attributes

**Discerning Speaker (DS)**
- Generate different attribute phrases from a specific object in an image
- They usually focus on the category, spatial relationships, etc., of the object in the image
- We focus on attributes that enable fine-grained discrimination with instances of a category

**Simple Speaker (SS)**
- Generate simple attribute phrases

**References**