

*Librería*  
***Bonilla y Asociados***  
*desde 1950*



**Título:**

**Autor:**

**Precio:** \$1000.00

**Editorial:**

**Año:** 2016

**Tema:**

**Edición:** 1ª

**Sinopsis**

**ISBN:** 9781608451128

Modeling data from visual and linguistic modalities together creates opportunities for better understanding of both, and supports many useful applications. Examples of dual visual-linguistic data includes images with keywords, video with narrative, and figures in documents. We consider two key task-driven themes: translating from one modality to another (e.g., inferring annotations for images) and understanding the data using all modalities, where one modality can help disambiguate information in another. The multiple modalities can either be essentially semantically redundant (e.g., keywords provided by a person looking at the image), or largely complementary (e.g., meta data such as the camera used). Redundancy and complementarity are two endpoints of a scale, and we observe that good performance on translation requires some redundancy, and that joint inference is most useful where some information is complementary.

Computational methods discussed are broadly organized into ones for simple keywords, ones going beyond keywords toward natural language, and ones considering sequential aspects of natural language. Methods for keywords are further organized based on localization of semantics, going from words about the scene taken as whole, to words that apply to specific parts of the scene, to relationships between parts. Methods going beyond keywords are organized by the linguistic roles that are learned, exploited, or generated. These include proper nouns, adjectives, spatial and comparative prepositions, and verbs. More recent developments in dealing with sequential structure include automated captioning of scenes and video, alignment of video and text, and automated answering of questions about scenes depicted in images