

# **Compositionality of MWEs : rich queries and priority of ranking documents**

**Su Nam Kim**

**CSSE dept. / University of Melbourne**

**September 21st, 2006**

**HCSNet Workshop on Next Generation Search**



**HCSNet**

The ARC Network in Human Communication Science

# Research Background (I)

- **Research Topic : Interpreting MWEs**
- **What is MWEs?**
  - **Def)** lexical items that can be decomposed into multiple simplex words
  - **Issues in MWEs**
    - Compound Nouns(*New York, apple pie*)
    - Verb-Particle Constructions(*battle on, put on*)
    - Light-Verb Constructions(*make a speech*)
    - non-decomposable idioms(*a piece of cake*)

# Research Background (II)

- **Issues to study**
  - Identification/Extraction of MWEs
  - Degree of Compositionality of MWEs
  - Interpretation of Semantics of MWEs
- **Approach**
  - Statistically modeling lexical acquisition tasks of MWEs using Similarity, Semantics and Grammatical Context

# Background

- **Focused Issues for IR**
  - MWEs as rich queries
  - MWEs to rank relevant documents
- **Aim to use MWEs in IR**
  - catch the users' intention in queries
  - provide more relevant documents in terms of various queries

# Motivation (I)

- **Enrich Queries**

- if multiple queries are compositional, it can be considered as **single query**
- e.g. cave art, take a walk, kick the bucket
- “cave”, “art” + “cave art”
- “take”, “walk” -> “walk”
- “kick”, “bucket” -> “die”

# Motivation (II)

- **Ranking Relevant Documents**

- When queries are highly compositional, priority of relevant documents by MWE queries should be higher
- w/ “a piece of cake”
- cf) in google, documents for food seem more relevant than documents for idiom, **“easy, simple”**

# Proposed Method (I)

- **w/ Compound nouns**
  - e.g. **cave art**
  - query : when multiple words as queries are highly compositional, then consider it as **single query**
  - rank : among retrieved documents, documents retrieved by “cave art” are **more relevant**
  - (cave art:289,000 vs cave procedure :37)

# Proposed Method (II)

- **w/ Light-Verb Constructions(LVCs)**
  - e.g. Mr. President **gave a speech**
  - query : w/ compositional queries as a LVC, use synonym(**speak**) of the LVC as a query (*give, speech* vs *speak*)
  - rank : documents retrieved by “speak” is not always relevant with the given queries\* i.e. one-way procedure)
  - *Mr.President spoke -/-> Mr.President gave a speech.*

# Proposed Method (III)

- **Non-decomposable Idioms**

- e.g. *a piece of cake*

- query : w/ mutiple words as an idiom, consider them as **single query**

- only one query from these keywords

- (*piece, cake vs a piece of cake*)

- rank : documents retrieved by MWE only have **relevant**

# Implications

- Handling MWEs provides enriched queries for IR
- MWEs as queries retrieve more accurate relevant documents
- MWEs provide broad range of search options