



Terminological Difference and Content Convergence: Automobile Patents from Mainland China and Taiwan

Benjamin K. Tsou

**Language Information Sciences Research Center,
City University of Hong Kong**

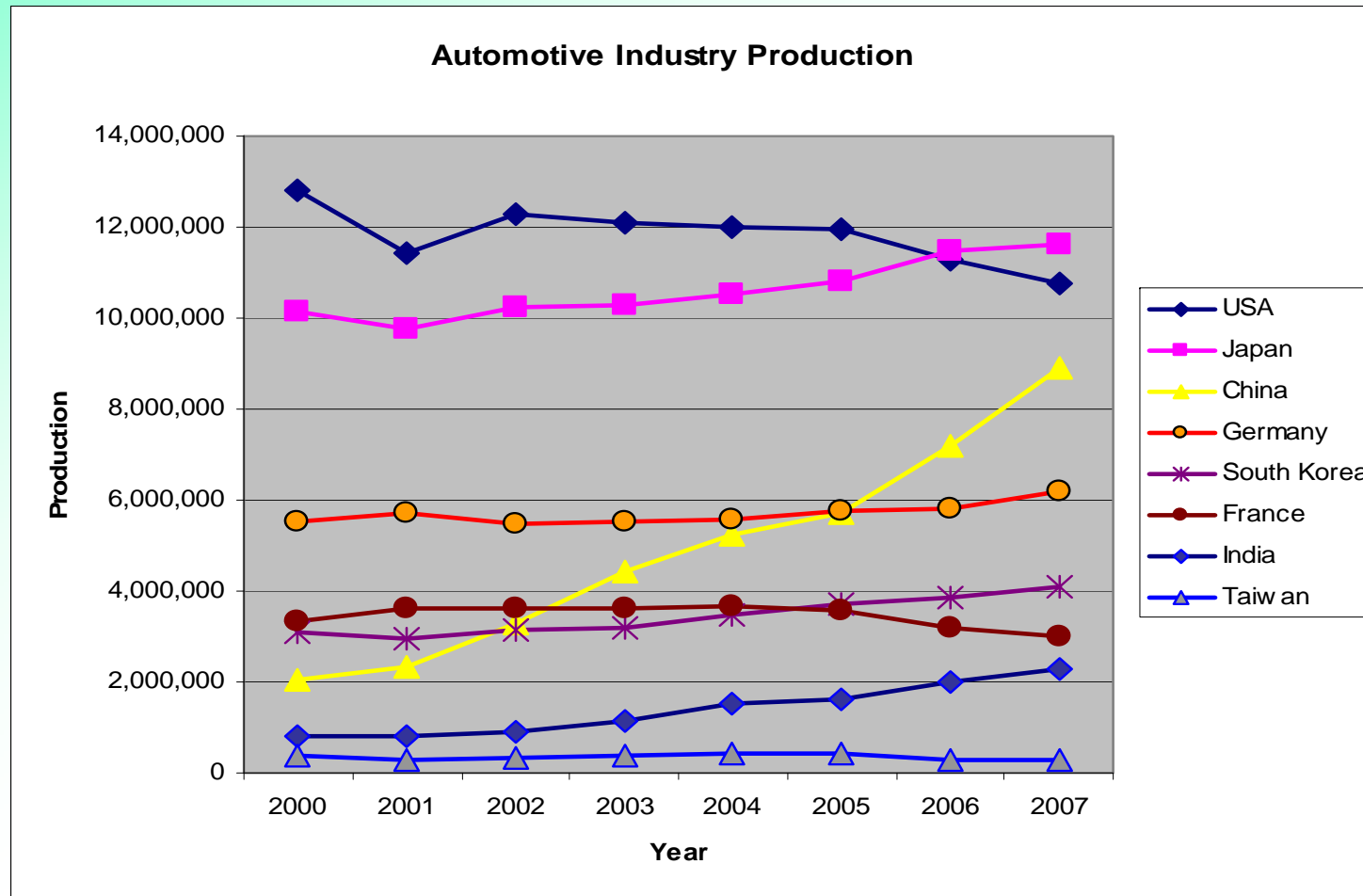


Introduction

- Automotive Patents from Mainland China and Taiwan
- Why automobile?
 - the importance of the automotive industry in the global economy
- Linguistic Diversity
 - Western languages and Chinese
 - Significant Pan-Chinese Variation
 - PRC, Taiwan, Hong Kong



Production Statistics of Automotive Industry

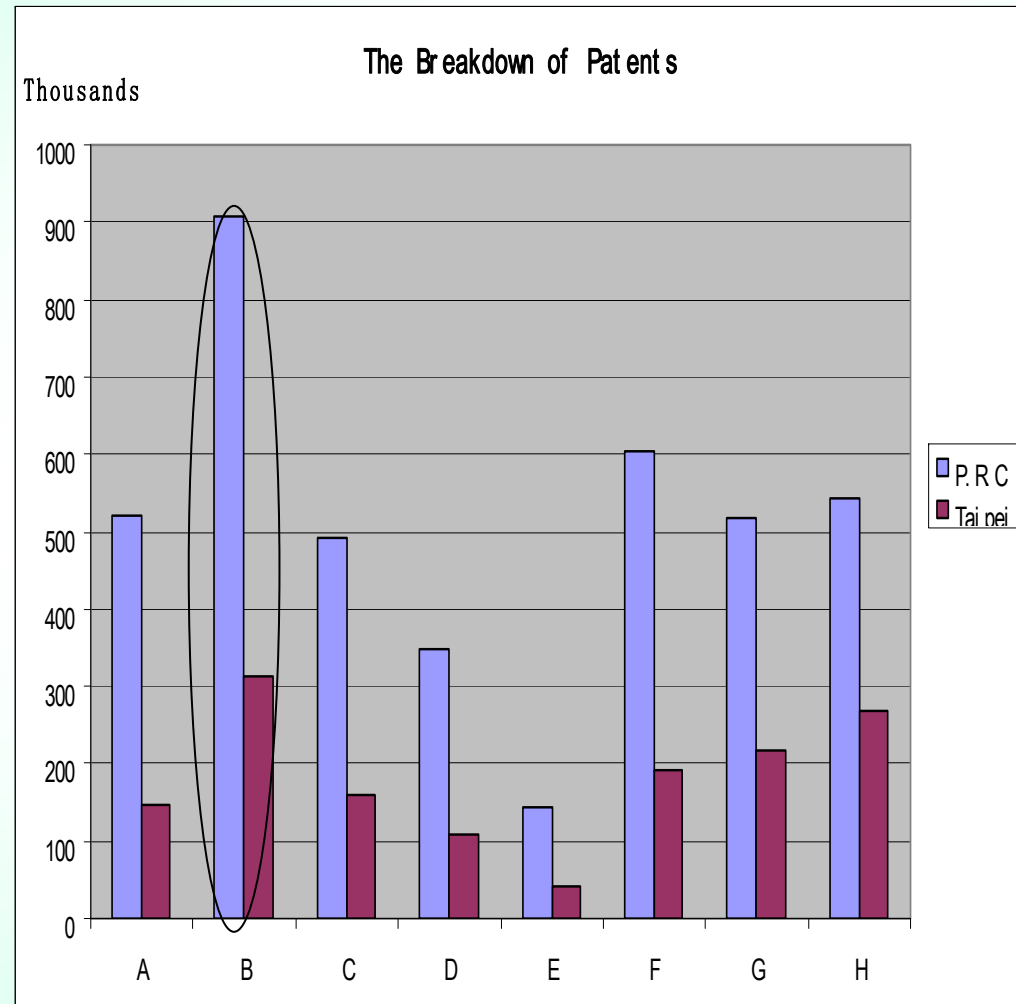




Total Number of Chinese Patents

IPC	Description
A	HUMAN NECESSITIES
B	PERFORMING OPERATIONS, TRANSPORTING
C	CHEMISTRY
D	TEXTILES
E	FIXED CONSTRUCTIONS
F	MECHANICAL ENGINEERING
G	PHYSICS
H	ELECTRICITY

Community	Count
P.R.C.	2,287,649
Taiwan	794,194
Hong Kong	123,442





IPC (International Patent Classification)

A: Human Necessities

B: Performing Operations, **Transporting**

Subsection: Transporting

– **B60 Vehicles in general**

– **B61 Railways**

– **B62 Land vehicles for travelling otherwise than on rails**

C: Chemistry, Metallurgy

D: Textiles, Paper

E: Fixed Constructions

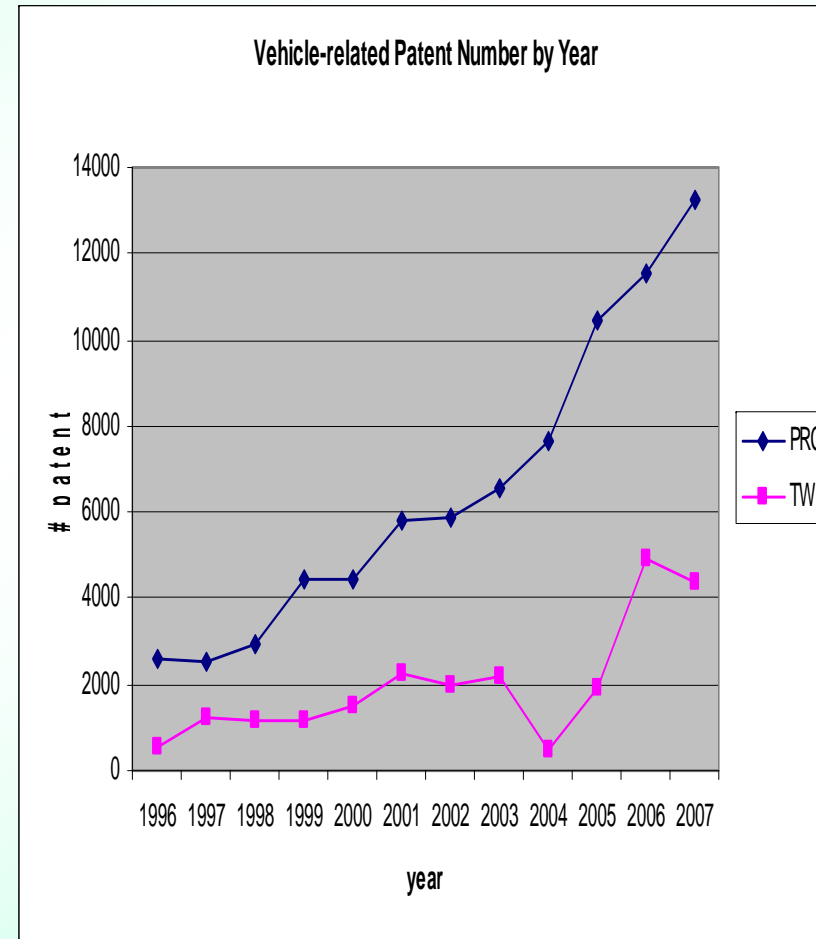
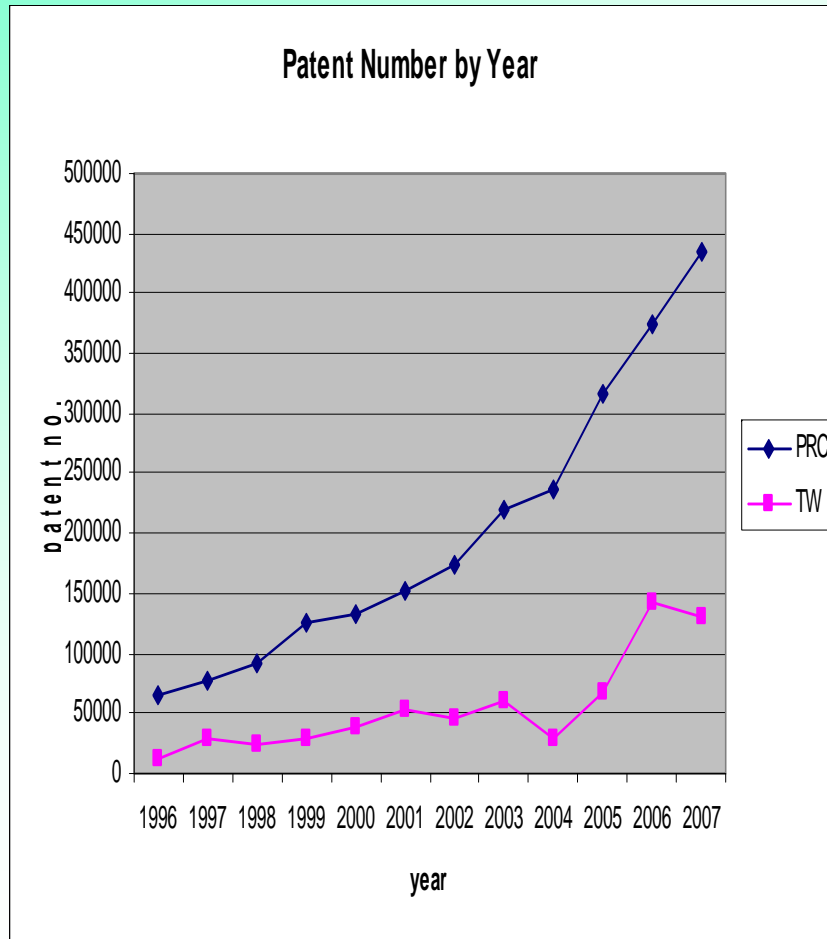
F: Mechanical Engineering, Lighting, Heating, Weapons,
Blasting

G: Physics

H: Electricity



Patent Number by Year





Patent Statistics for PRC and Taipei (2008 Summer)

Area	Total # of Patents	Total # of Auto.-related Patents	Total # for IPC B60	Total # for IPC B61	Total # for IPC B62
PRC	2,322,414	108,125	62,307	9108	36710
Taipei	800,850	36,008	19,003	370	17925



a Corpus of Chinese Vehicle-related Patents

	PRC	TW
# of Patents	95,307	34,420
Avg. Num. of Characters Per Patent	Abstract: 220 Claim: 230	Abstract: 230 Claim: 310
Corpus Size (Characters)	Abstract: 20M Claim: 21M	Abstract: 7M Claim: 10M

Notes:

1. The claims of PRC's patents extracted from web pages are only a small part of the whole claims.
2. Only some Taiwan's patents have abstracts and claims.



1. Analysis of Parallel Patents

- Parallel patents
 - Over 400 linguistically paired (*LP*) patents expected
 - A set of *LP* patents belong to a patent family, resulting from related patent applications in different languages
 - English equivalents from the web site of WIPO.

	Sentences	Characters	Words
Mainland	1,828	179,720	107,631
Taiwan	2,133	173,176	99,433

Table 1. Some Examples



Parallel lang Patent Sample 1

Title:

EN: EMERGENCY PNEUMATIC SPRING WITH CENTRING ACTION

PRC: 具對中作用的緊急充氣彈簧

TW: 自動定心之緊急彈簧支承

Abstract:

EN: The invention relates to a *pneumatic spring* (LUF) for a rail vehicle (SFZ). *Said spring comprises* two *bellows rims* (BF1, BF2), mounted one above the other and interconnected by *pneumatic-suspension bellows* (LBA) ...

PRC: 本發明關於一用於軌道車輛(SFZ)的充氣彈簧(LUF)。該彈簧包含兩個囊架(BF1, BF2)，其中一個裝設於另一個上面，並由充氣懸置氣囊(LBA)連接的...

TW: 本發明係關於一種軌道車輛(SFZ)用之空氣彈簧(LUF)，其具有二於組裝狀態時，上下設置且藉由氣囊(LBA)相連接之囊架(BF1、BF2) ...



Patent Sample 4

Title:

EN: System and device for locking an **automobile steering axis**

PRC: 用於鎖定**汽車轉向軸**的系統和裝置

TW: **車輛駕駛軸**鎖定系統及裝置

Abstract:

EN: A system and device for locking an **automobile steering axis** according to the present invention is an **automobile** stealing prevention device permanently installed in an automobile provided with of an electrical-electronic system to control the **lock and unlock** mechanism of the **automobile steering axis** to facilitate locking operation....

PRC: 根據本發明的用於鎖定**汽車轉向軸**的系統和裝置是一種永久性地安裝在汽車上的**汽車**防盜裝置，該裝置設置有電氣電子系統，以控制**汽車轉向軸**的**鎖定和解鎖**機構，使鎖定操作變得更加方便。...

TW: 本發明之**車輛駕駛軸**鎖定系統及裝置係一常駐車內之**車輛**防盜裝置，其具有一電氣-電子系統用以控制**車輛駕駛軸**機構之**閉鎖與開鎖**操作。...



Term Comparison between Mainland and Taiwan

English	Mainland	Taiwan
Interface	接口	介面
multi-axle	多軸	多輪軸
A brake system	制動系統	煞車系統
configuration information	配置信息	組態資訊
sensor	傳感器	感測器
data	數據	資料
information	信息	資訊

Reasons for Term Differences

1. Two writing systems:
2. limited communication between mainland China and Taiwan on patent matters



2. Analysis of Comparable Patents

- Randomly selected patents from CityU corpus
 - 100 PRC patents and 112 Taiwan patents.
 - Patents manually checked for relevance to automobile.
- Abstracts and claims
 - Word Segmentation and POS-tagging
 - using the very large LIVAC dictionary and some rules
 - Results are manually verified.



What is the overall number of
automotive technical terms,
not in 1.2 million lexical database of
“ordinary language” i.e.
LIVAC (<http://livac.org>)?



Analysis of Technical Terms

- Technical Term Extraction
 - Compare LIVAC dictionary (below)
 - **Nouns**
- Results (manually check)
 - **Large amount of technical terms** (35% of nouns)
 - only in patents, but not in LIVAC (12 years)
 - **Precision:** Almost 100%
 - all nouns not appearing in LIVAC are **technical terms**
 - **Recall:** About 89.8% (LIVAC 06-07)
 - a small portion (10.2%, 93/909) also appears in LIVAC (1 years)

	# of patents	# of noun types	Overlapping with LIVAC 06-07 (1 year)	Overlapping with LIVAC 02-07 (5 years)	Overlapping with LIVAC 95-07 (12 years)
PRC	100	1567	816 (52.1%)	948 (60.5%)	1017 (64.9%)
Taiwan	112	1552	696 (44.8%)	846 (54.5%)	890(57.3%)



Similarity Analysis between Technical Terms

	A. # of noun types	B. identical	A (%)	B/A (%)
1 identical character	443	92	58.99	20.77
2 identical characters	181	91	24.10	50.28
3 identical characters	34	31	4.53	91.18
No identical characters	93	0	12.38	0
Total number	751	214	100	28.50



Challenges Ahead

- Domain-specific New Word Detection in Chinese
- Detecting corresponding words in PRC patents and Taiwan patents with different methods:
 1. Word alignment based on parallel patents
 2. Logograph-based approach
 3. Domain-ontology-based approach
 4. Word co-occurrence probabilities in texts
 5. **Comparative neological development



Method 1:

Word Alignment based on parallel patents

- Extraction of technical terms
 - Sentence alignment
 - Word alignment
 - Statistical and heuristic alignment models

EN: A *brake* system for a *multi-axle* vehicle is described *which includes a plurality of brake elements* for applying a required braking force to the vehicle in response to a brake command and a controller which ...

PRC: 公開了一種用於多軸車輛的制動系統，其包括多個制動元件和控制器，制動元件響應制動命令而向該車輛施加所需制動力...

TW: 本發明提出一種用於多輪軸車輛之煞車系統，該煞車系統包含回應於一煞車命令而對該車輛施予一所需制動力之複數個煞車元件...



Method 2:

Logographic-based approach

The *hypothesis*:

Most corresponding words in comparable PRC and Taiwan patents have some identical characters.

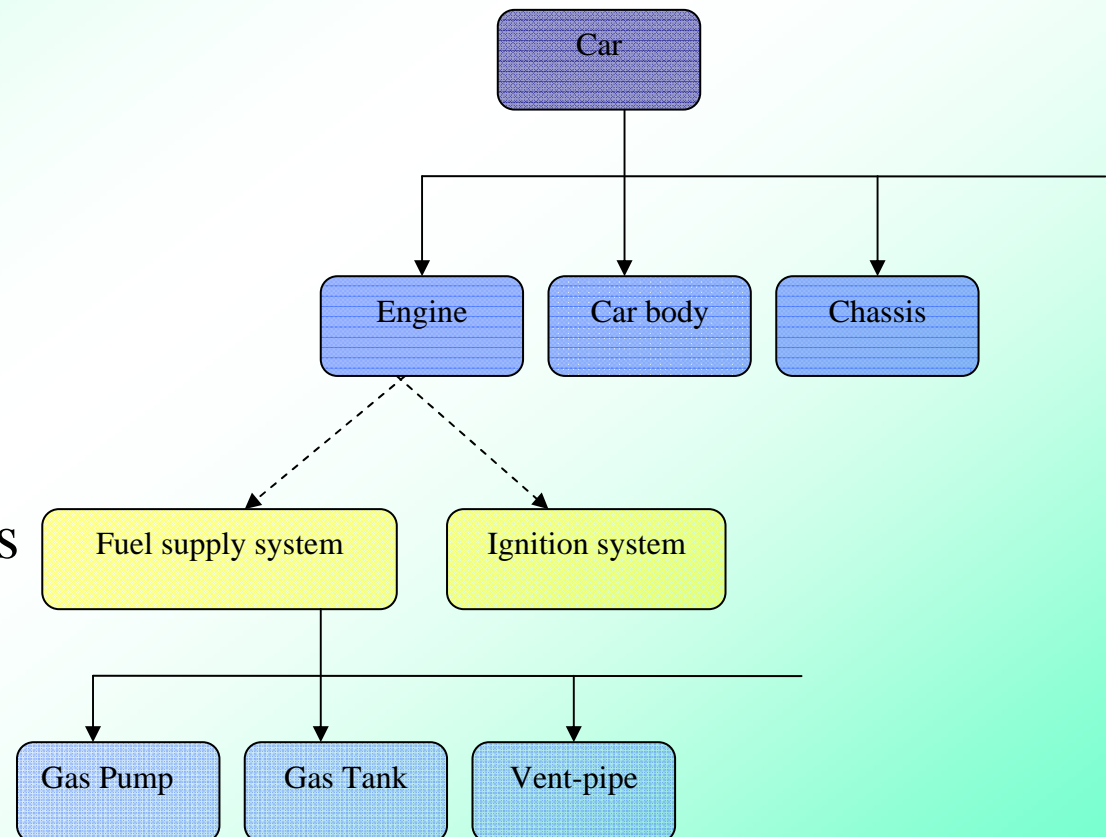
PRC's Noun Types	Taiwan's Noun Types	# Identical Characters	English Explanation
橫斷面	剖面	1	Section plane
氣缸	汽壓缸	1	Cylinder
報信器	蜂鳴器	1	Notifying device
變換器	轉換器	2	Transducer
點火器	點煙器	2	Igniter
散熱器	散熱器	3	Radiator
釋放桿	釋放桿	3	Release lever



Method 3:

Domain-ontology-based approach

- IPC categories
 - Comparative Patents belonging to the same category
- Technical Terms
 - To use categories appearing these terms as features
 - To compute their similarity based on features





Method 4: Word Co-occurrence Probability

- Extract correspondent technical terms from comparable corpora and an initial bilingual lexicon.
- Compare the distributional contexts of source and target words
 - testing several weighting factors and similarity measures.



Word Statistics

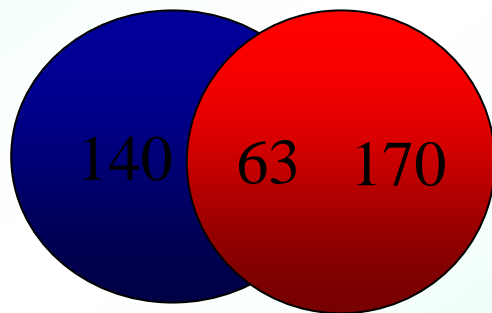
for 20 parallel patent abstracts

	# word tokens	# average word token	# word types
PRC	2234	112	643
Taipei	2945	147	794
English	3184	159	653
Combined (PRC&Taipei)	5809	290	1119

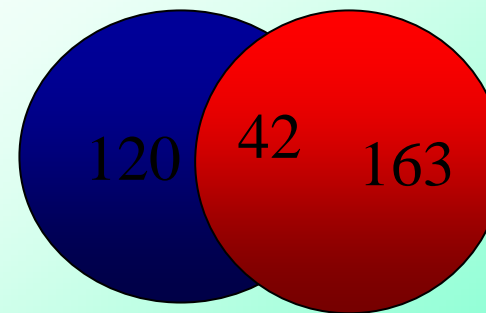


New Terms Statistics for 20 parallel patent abstracts

	# word types not in LIVAC06-7	# special terms related to auto.	# noun types not in LIVAC06-07	# new special nouns related to auto.
PRC	185	140(75.7%)	120	107(89.2%)
Taipei	249	170(68.3%)	163	145(89.0%)
Combined (PRC&Taipei)	362	247(68.2%)	241	217(90.0%)



1. Special Terms Overlapping



2. Special Nouns Overlapping



Conclusion

- A Preliminary Analysis of a Corpus of Chinese Automobile-related Patents indicates unusual challenges because of critical linguistic variation
- Future work
 - determine differences between Mainland China and Taiwan in neological creations and use of automotive related terminologies.
 - explore the methodological basis for construction of parallel corpus of Chinese language automotive patent files in Mainland China and Taiwan
 - cross-lingual correspondences and retrieval between English and Chinese



Thanks!

Q & A



- England and America are”divided by a common language”