

A corpus-based study of the three connectives
danshi , *keshi* , and *ran'er*
in Mandarin Chinese

American Association for Corpus Linguistics(AACL)
October 11, 2009
Ruiting Jia
University of Alberta

Introduction

■ *danshi, keshi, ran'er* in Mandarin Chinese

- Equivalent to *but, however* in English
- Previous studies (Lü 1979, Li & Thompson 1981):
 - (1) Considered to be synonymous
 - (2) Contrast/Contradiction
 - (3) Favored in clause-initial position

Introduction

■ *danshi, keshi, ran'er* in Mandarin Chinese

- Equivalent to *but, however* in English
- Previous studies (Lü 1979, Li & Thompson 1981):
 - (1) Considered to be synonymous
 - (2) Contrast/Contradiction
 - (3) Favored in clause-initial position

(1) *Jintian zaoshang benlai yao kaihui, danshi linshi quxiao le.*
Today morning actually have meeting, but temporally cancel particle.
'There was a meeting scheduled for this morning, but it's been canceled'

(2) *Ta yinggai jintian lai xuexiao de, keshi ta wang le.*
He should today come school particle, but he forgot particle.
'He should come to school today, but he forgot'

(3) *Ta xiang shi mei wenti de, ran'er haishi chu le wenti.*
He think is no problem particle, but still have particle problem.
'He thought there would be no problem, but it still have the problem.'

(4) *keshi*: often appears in clause initial position, but can also be immediately preceded by a noun phrase (Li & Thompson 1981).

(4) *keshi*: often appears in clause initial position, but can also be immediately preceded by a noun phrase.

(5) Usage:

keshi: mainly used in **Spoken** Mandarin

ran'er: mainly used in **Written** Mandarin

(4) *keshi*: often appears in clause initial position, but can also be immediately preceded by a noun phrase.

(5) Usage:

keshi: mainly used in **Spoken** Mandarin

ran'er: mainly used in **Written** Mandarin

But.....

Questions

- ▶ Frequency distribution

- (1) Across in different genres

- (2) In clause-initial & non-initial position in both written and spoken context

- ▶ Semantic associations

- (3) Negative & Positive attitude of speaker/writer

Current corpus-based study

▶ Methods:

- Frequency distribution: statistics analysis
- Semantic association: concordance lines:

▶ Two corpora:

- **LCMC_w**: Mandarin **written** corpora (15 genres/1,000,000 words)
- **Callhomes**: Mandarin **spoken** corpora (120 files/300,767 words)

▶ Corpus tool:

Wordsmith and Xaira

Questions

- ▶ Frequency distribution

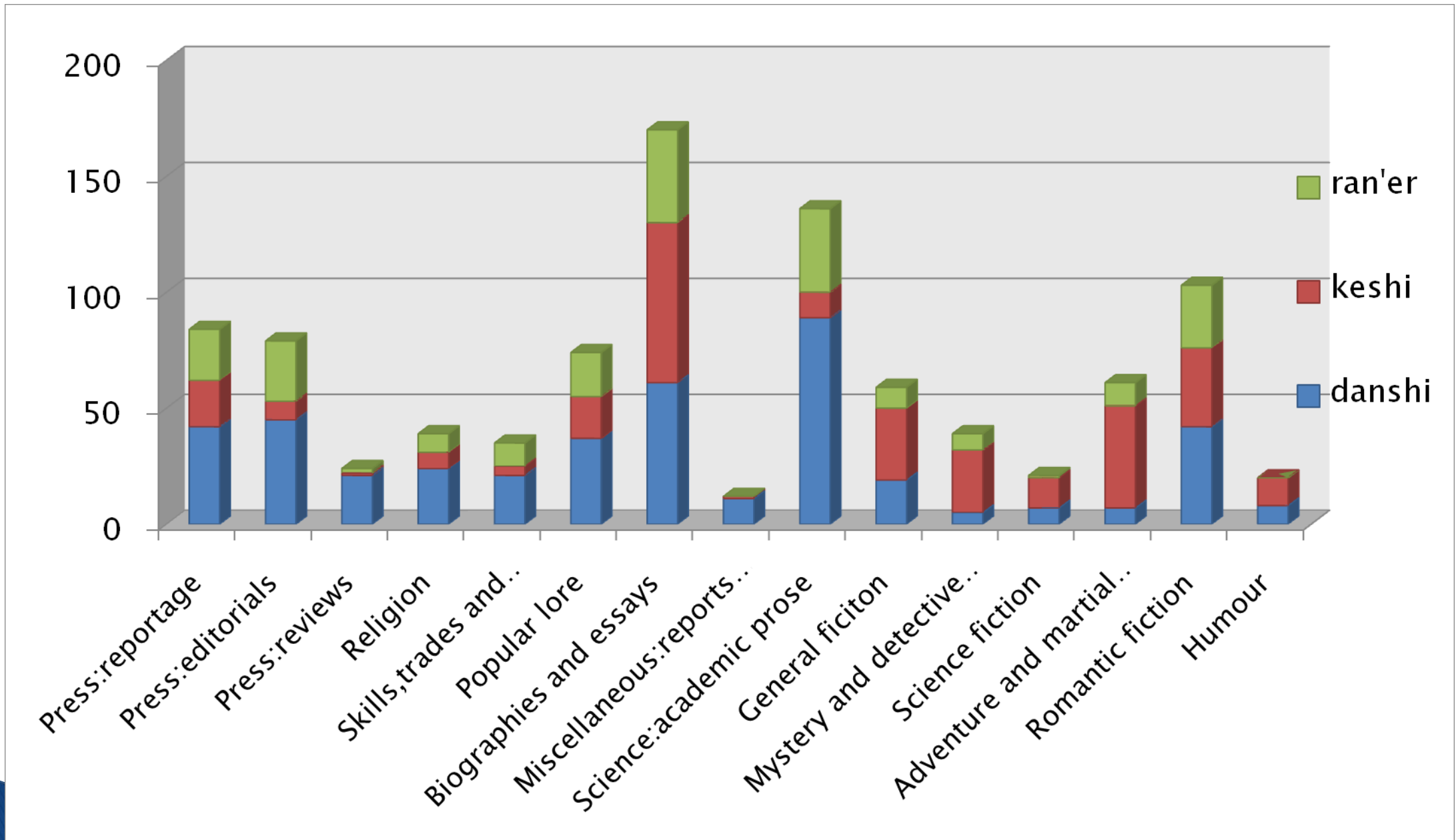
- (1) Across in different genres

- (2) In clause-initial & non-initial position in both written and spoken context

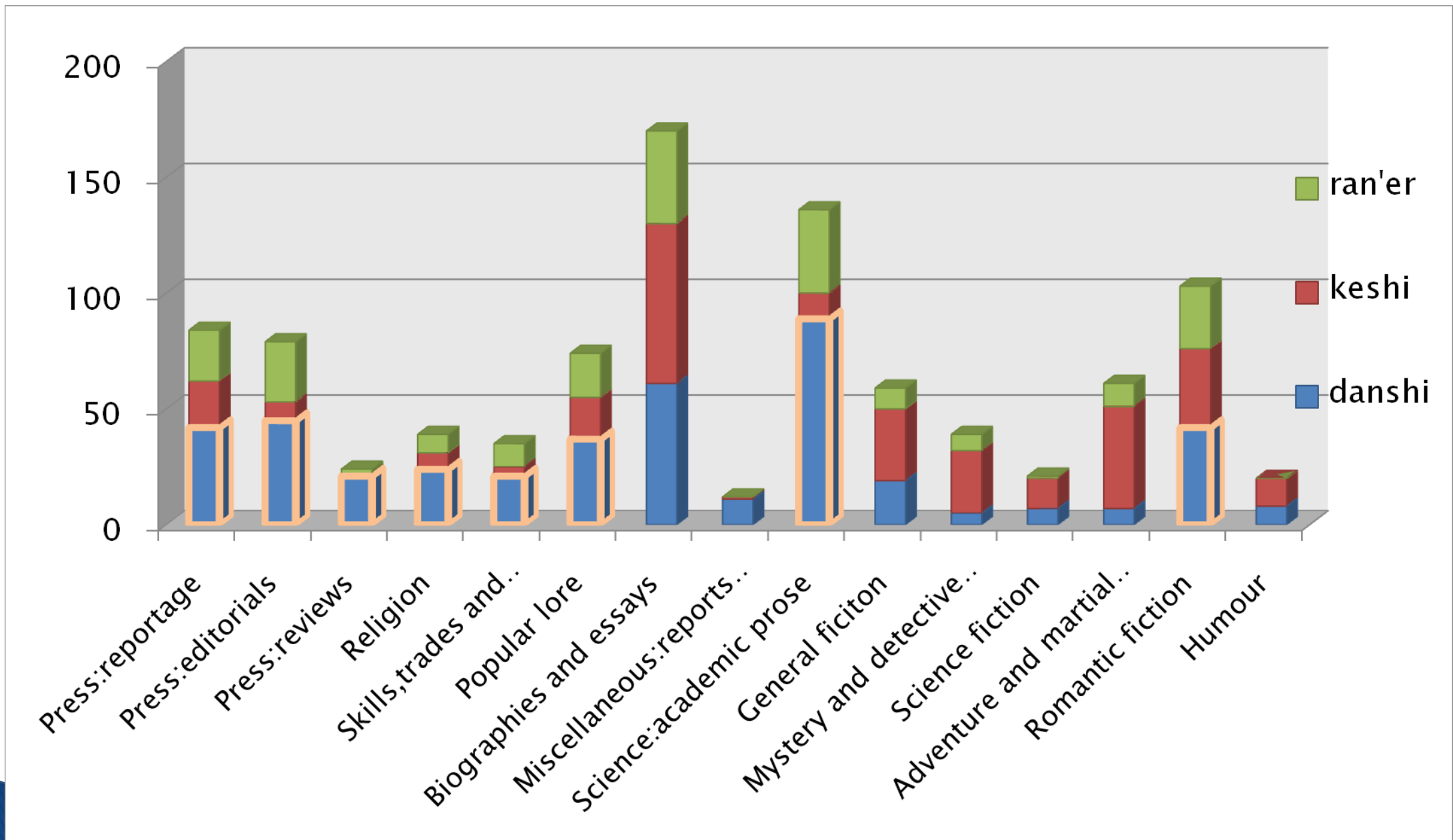
- ▶ Semantic associations

- (3) Negative & Positive attitude of speaker/writer

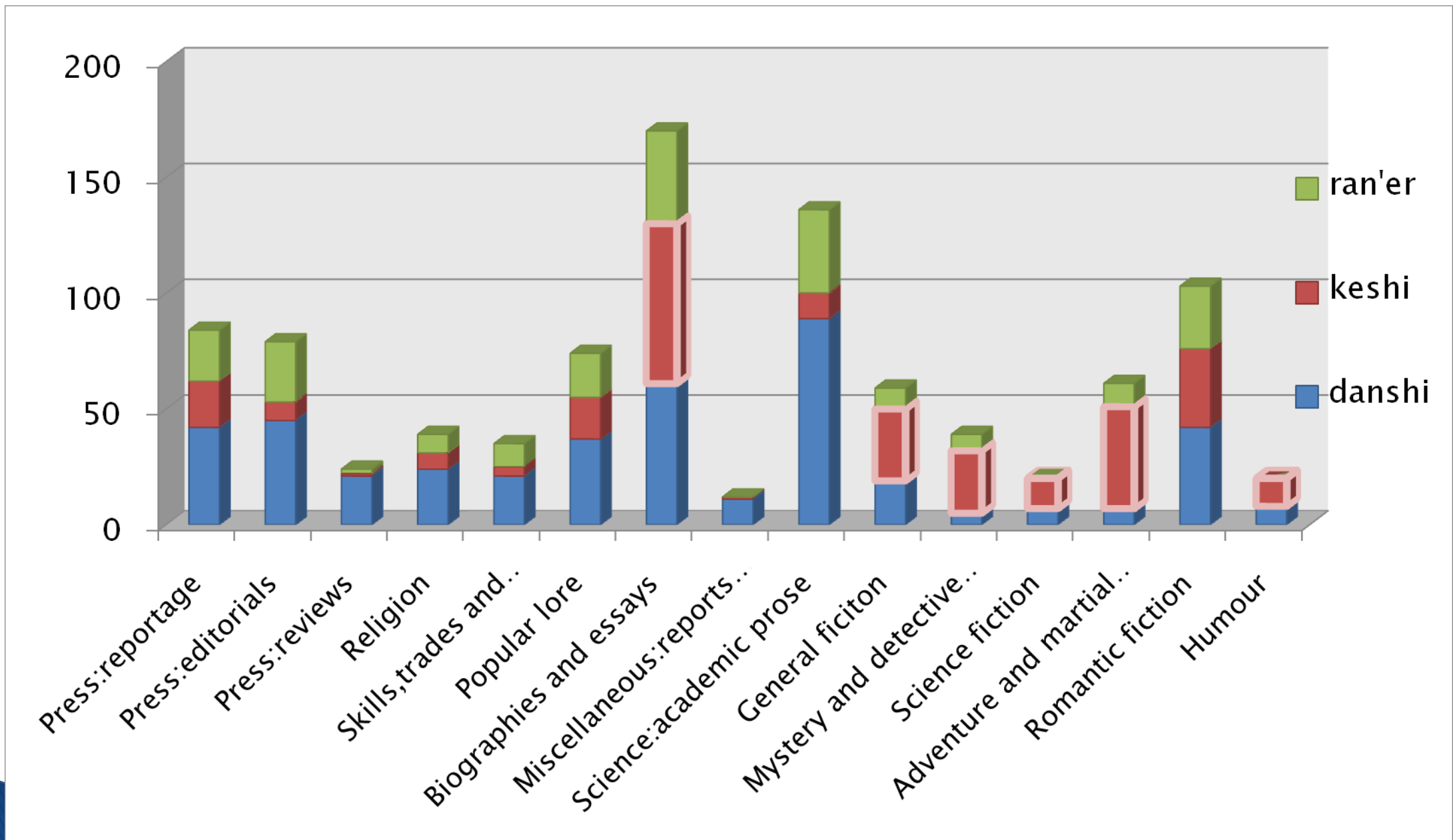
Frequency distribution of the three connectives in LCMC_w



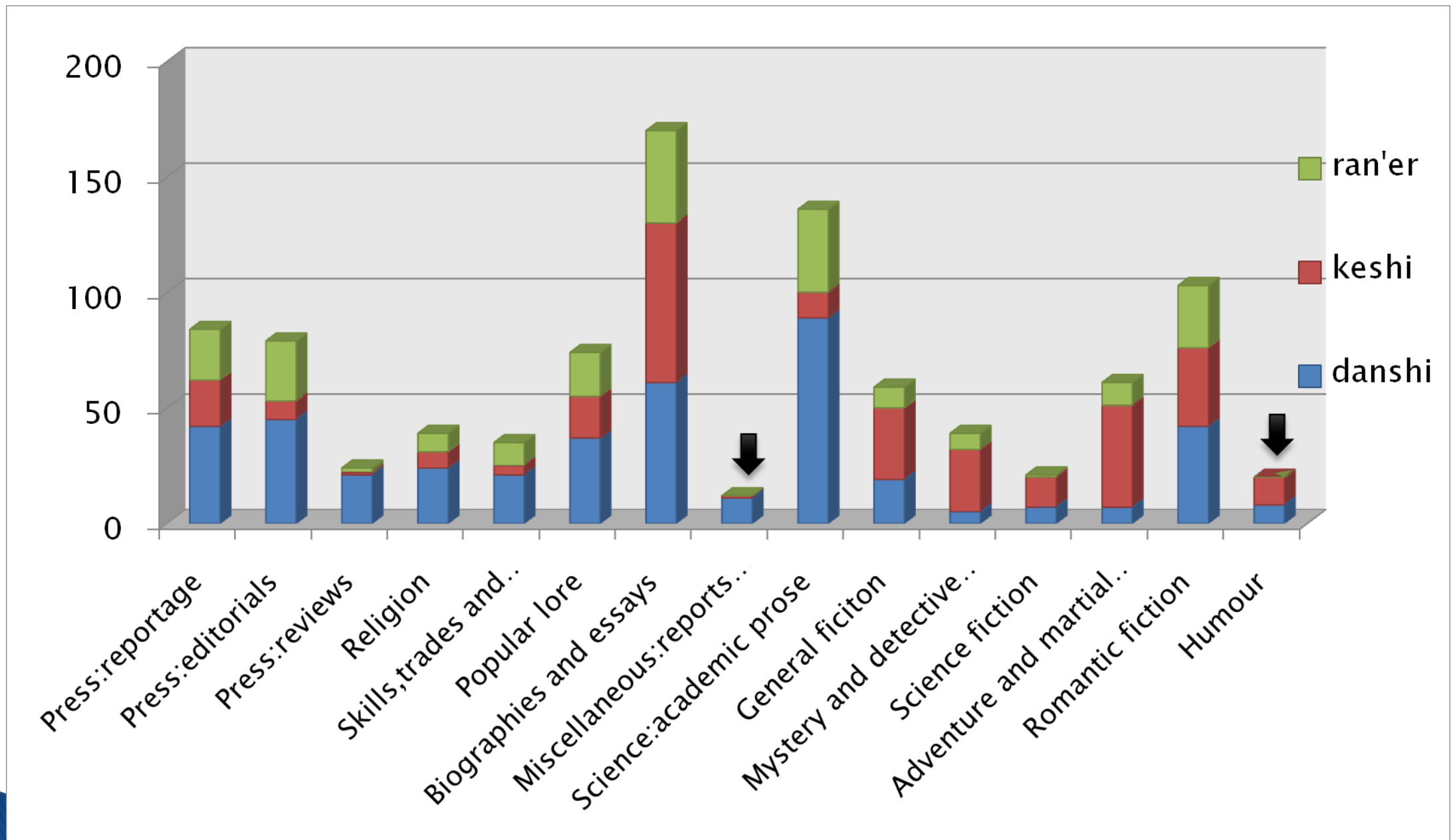
Frequency distribution of the three connectives in LCMC_w



Frequency distribution of the three connectives in LCMC_w



Frequency distribution of the three connectives in LCMC_w



Frequency distribution of the three connectives in LCMC_w & Callhomes

Connective	LCMC _w Total/genre	Callhomes Total/file
danshi	435/15	403/96
keshi	302/15	52/32
ran'er	196/13	3/3

Frequency distribution of the three connectives in LCMC_w & Callhomes

Connective	LCMC _w Total/genre	Callhomes Total/file
danshi	435 /15	403 /96
keshi	302/15	52/32
ran'er	196/13	3/3

Frequency distribution of the three connectives in LCMC_w & Callhomes

Connective	LCMC _w Total/genre	Callhomes Total/file
danshi	435/15	403 /96
keshi	302/15	52 /32
ran'er	196/13	3/3

Frequency distribution of the three connectives in LCMC_w & Callhomes

Connective	LCMC _w Total/genre	Callhomes Total/file
danshi	435/15	403/96
keshi	302/15	52/32
ran'er	196/13	3/3

Questions

- ▶ Frequency distribution

- (1) Across in different genres

- (2) In clause-initial & non-initial position in both written and spoken context

- ▶ Semantic associations

- (3) Negative & Positive attitude of speaker/writer

Positional distribution of the three connectives

Connective	LCMC _w		Callhomes	
	Initial	Non-initial	Initial	Non-initial
danshi	99%	1%	92%	8%
keshi	87%	13%	85%	15%
ran'er	99%	1%	100%	0%

Positional distribution of the three connectives

Connective	LCMC _w		Callhomes	
	Initial	Non-initial	Initial	Non-initial
danshi	99%	1%	92%	8%
keshi	87%	13%	85%	15%
ran'er	99%	1%	100%	0%

Positional distribution of the three connectives

Connective	LCMC _w		Callhomes	
	Initial	Non-initial	Initial	Non-initial
danshi	99%	1%	92%	8%
keshi	87%	13%	85%	15%
ran'er	99%	1%	100%	0%

Positional distribution of the three connectives

Connective	LCMC _w		Callhomes	
	Initial	Non-initial	Initial	Non-initial
danshi	99%	1%	92%	8%
keshi	87%	13%	85%	15%
ran'er	99%	1%	100%	0%

Positional distribution of the three connectives

Connective	LCMC _w		Callhomes	
	Initial	Non-initial	Initial	Non-initial
danshi	99%	1%	92%	8%
keshi	87%	13%	85%	15%
ran'er	99%	1%	100%	0%

Non-initial position: examples from LCMC_w

- *danshi*

(1) adverb + **danshi**

....., *disheng danshi xiangdang zhi yansu*.....

....., *lowly but fairly particle seriously*

- *ran'er*

(2) verb + **ran'er**

....., *jiechuan ran'er zhaoyang*

....., *unveil but still*

- *keshi*

(1) Noun/pronoun + **keshi**

....., *zuichun keshi hen hou*.....

....., *lips but very thick*.....

(2) adverb + **keshi**

....., *xianzai keshi tichang*.....

....., *currently but promote*.....

Non-initial position: examples from Callhomes

- *danshi*

(1) Repetition

....., *danshi danshi* *wo de dao de xiaoxi feichang shao*.....

....., *but but my get particle message very few*.....

(2) Particle + *danshi*

....., *e danshi* *wo shuo ni*

....., *uh but I say you*.....

(3) Connective + *danshi*

....., *yinwei danshi na ge ren*.....

....., *because but that person*.....

- *keshi*

(1) Repetition

(2) Modal verb + *keshi*

....., *yingai keshi mang de*

....., *should keshi busy particle*

Questions

- ▶ Frequency distribution

- (1) Across in different genres

- (2) In clause-initial & non-initial position in both written and spoken context

- ▶ Semantic associations

- (3) Negative & Positive attitude of speaker/writer

Distribution of three connectives across different semantic associations in LCMC_w & Callhoms

Connective	LCMC _w			Callhoms		
	Positive	Neutral	Negative	Positive	Neutral	Negative
danshi	39%	6%	55%	17%	40%	43%
keshi	24%	6%	70%	8%	50%	42%
ran'er	24%	3%	73%	0%	67%	33%

Distribution of three connectives across different semantic associations in LCMC_w & Callhoms

Connective	LCMC _w			Callhoms		
	Positive	Neutral	Negative	Positive	Neutral	Negative
danshi	39%	6%	55%	17%	40%	43%
keshi	24%	6%	70%	8%	50%	42%
ran'er	24%	3%	73%	0%	67%	33%

Distribution of three connectives across different semantic prosodies in LCMC_w & Callhoms

Connective	LCMC _w			Callhoms		
	Positive	Neutral	Negative	Positive	Neutral	Negative
danshi	39%	6%	55%	17%	40%	43%
keshi	24%	6%	70%	8%	50%	42%
ran'er	24%	3%	73%	0%	67%	33%

Examples from LCMC_w & Callhomes

▶ *danshi*

(1) ... *tongshi chansheng, danshi shuliang jiao shao ...*
... *meantime occur, but number less ...*

(2) ...*danshi ta zhongyu xue hui le youyong ...*
...*but he finally learn able particle swim ...*

(3) ...*danshi keji de jinbu sudu yuanyuan chaoguo ...*
...*but technology particle progress speed far beyond ...*

Examples from LCMC_w & Callhomes

▶ *keshi*

(1) ..., *keshi* *shenm* *ye* *mei* *kandao*....
... *but* *nothing* *either* *not* *seen*...

(2) ..., *keshi* *youyu* *chuantong* *wenhua* *de* *yingxiang*...
..., *but* *because* *traditional* *cultural* *particle* *influence*...

(3) ... , *keshi* *shidai* *jinbu* *le*...
..., *but* *now* *improve* *particle*...

Examples from LCMC_w & Callhomes

▶ *ran'er*

(1) ..., *ran'er* women **cuo** le.

..., *but* we **wrong** particle.

(2) ..., *ran'er* zhongyu zhi you huiqu le.

..., *but* finally closed friend back particle.

(3) ..., *ran'er* ni you yi ge **zui youyong** de pengyou...

..., *but* you have one classifier **most useful** particle friend...

Conclusion

- ▶ **Frequency distribution**
 - *ran'er*: mainly used in written

Conclusion

- ▶ **Frequency distribution**

- *ran'er*: mainly used in written



Conclusion

- ▶ **Frequency distribution**

- *ran'er*: mainly used in written



- *keshi*: mainly used in spoken

Conclusion

▶ Frequency distribution

- *ran'er*: mainly used in written



- *keshi*: mainly used in spoken



Conclusion

▶ Frequency distribution

- *ran'er*: mainly used in written



- *keshi*: mainly used in spoken



- Favor in clause-initial position

Conclusion

▶ Frequency distribution

- *ran'er*: mainly used in written



- *keshi*: mainly used in spoken



- Favor in clause-initial position



Conclusion

▶ **Semantic Prosody:**

Three connectives associate with three different semantic prosodies. The most predominant semantic association in **written** Mandarin is more likely...

Conclusion

▶ **Semantic Prosody:**

Three connectives associate with three different semantic prosodies. The most predominant semantic association in **written** Mandarin is more likely...

Negative

Implications

- ▶ Formal grammar teaching: L1 & L2
 - Patterns
 - Contexts
 - Semantic association

谢谢
XIEXIE
THANK YOU