Grammaticality and Parsability in Sentence Processing

Chien-Jer Lin
National Taiwan Normal University
clin@ntnu.edu.tw

International Workshop on Grammar and Evidence, Chiayi, April 13-15 2007

acknowledgements

- James Myers for stimulating thoughts on the issue between grammar and evidence
- NSC grant (NSC 95-2411-H-003 -056)
- Research Assistants: Li-Hsin Ning, Larry Li
- NTNU research grant
- Yu-Bin Chen (NCCU) and Ray Huang (NTNU) for providing some of the RVC data + inspirations
- Tom Bever & Heidi Harley for numerous discussions; Thomas Lee for insights on Chinese garden-path sentences

Typical questions syntacticians ask of their friends:

- Does this sentence sound good (I mean, grammatical) to you?
  這個句子聽起來可以嗎？
  - Hmm, sounds weird. (after repeating several times) maybe it's alright. But, I wouldn't say so myself.
    好像可以，但是我不會這樣講。
    聽起來怪怪的，哪裡找來的句子，像是英文翻譯過來的。我們中文不會這麼說吧。

Typical questions syntacticians ask of their friends:

- How about this one? Do you like it better?
  那這句呢？你真的不喜歡嗎？
  - Are they different? No, still sounds funny.
    還是怪怪的呀，可能可以吧，這種句子沒有人會講。

Bever (1970)

"The influence of speech performance on linguistic structure."

- "It is simultaneously the greatest virtue and failing of linguistic theory that sequence acceptability judgments are used as the basic data."

Bever (1970)

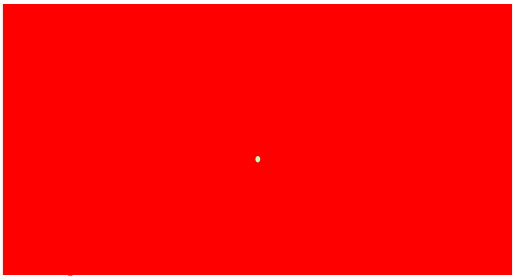
"The cognitive basis for linguistic structures."

Important factors in accessing our linguistic competence:

- Perceptual strategies in general cognition.
- Perceptual strategies in language—NVN

Bever (1970): Perceptual illusions can be misleading.

Bever (1970): Perceptual illusions can misleading.



T. Langendoen (1972) “The Problem of Grammaticality.”

***Peabody Journal of Education* 50, 20-23.**

“Grammar is in trouble and needs help. (20)”

“Salience of one interpretation [can be] so much greater than the other, that the other is totally suppressed, perceptually. (22)”

9

T. Langendoen (1972) “The Problem of Grammaticality.”

***Peabody Journal of Education* 50, 20-23.**

“Raw grammaticality judgments cannot be interpreted always at face value; that to obtain “true” grammaticality judgments one must at least factor out effects that are the results of the inner workings of rhetoric and logic (23)”

10

F. Ferreira (2005) “Psycholinguistics, Formal Grammars, and Cognitive Science.” *Linguistic Review*.

- Suggestions on grammatical judgments:
  - Generative grammarians should rely on **more than grammaticality judgments** when developing theories of structure.
  - The **methods** of collecting judgments should be improved.

11

Lessons we should have learned in the past 40 years.

- Grammar resides in human beings who are finite in their ability to cognize.
- Linguists should pay attention to the potential “**cognitive**” and/or “**perceptual**” factors in grammaticality judgments so that grammar can be minimized.

12

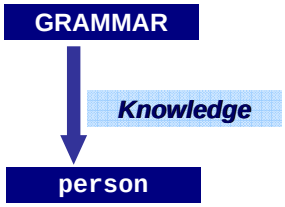
Chomsky (1981) "Knowledge of Language: Its Elements and Origins."

- **Generative Grammar**
- "**Knowledge** of Language"—
  - The nature of this knowledge
  - How is it acquired?
  - How is it used?

Conceptual Priority

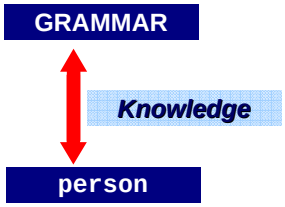
13

Chomsky (1981) "Knowledge of Language: Its Elements and Origins."



14

Chomsky (1981) "Knowledge of Language: Its Elements and Origins."



15

Chomsky (1981) "Knowledge of Language: Its Elements and Origins."

- "Linguistics is a part of psychology, and ultimately biology. (23)"

16

Issues

- Factors that affect grammaticality judgments
- **parsability**
- Are ungrammatical sentences actually difficult sentences?
  1. Processing of relative clauses with nested dependencies
  2. Thematic mappings of Mandarin RVCs (Resultative Verb Compounds)

17

The one thing that psychologists care about most in their methodology

Confounds

18

The goal of sentence processing research

- to figure out how grammaticality among other factors affect the way we understand sentences

19

Factors that affect grammaticality judgments

20

Factors that affect grammaticality judgments (Schütze, 1996)

- **Subject-related** factors (Individual differences)
  - Capacity of working memory
  - Handedness
  - Linguistic training
  - Literacy and education

21

Factors that affect the goodness of a sentence (Schütze, 1996)

- **Task-related** factors
  - Instructions
  - Order of presentation
  - Repetition
  - Mental state
  - Judgment strategy
  - Modality and register
  - Speed

22

Factors that affect the goodness of a sentence (Schütze, 1996)

- **Task-related** factors
  - **Stimuli**
    - Context
    - Meaning
    - **Parsability**
    - Frequency
    - Lexical content
    - Morphology and spelling
    - Rhetorical structure

23

Parsability as the basis for some grammaticality judgments

- Some sentences can be so difficult that subjects decide that they are not possible (to understand?).
- Subjects confuse implausibility or difficulty with ungrammaticality and decide that certain sentences are ungrammatical.

24

Parsability

- Sentences that take a lot of WM to parse.
- Sentences that are misparsed and a correct parse was never reached.

25

1. Sentences beyond normal working memory capacity

- One classic example has been the processing difficulty with nested dependencies.

26

1. Sentences beyond normal working memory capacity

- For instance:

27

Limited working memory for sentence processing

- Center embeddings are difficult. (Bever, 1970; Miller & Chomsky, 1963; Gibson, 1998)

The pupil [the proctor spotted] cheated.

28

1. Sentences beyond normal working memory capacity

- In this section, I demonstrate the limit of working memory and the effect of sentence parsability on the accuracy of comprehension.

29

1. Sentences beyond normal working memory capacity

- Experiment 1—Self-paced reading of sentences with Mandarin relative clauses (Lin, 2006 & 2007)
  - 1a: sentences with single RCs
  - 1b: sentences with double RCs

30

Some properties of Mandarin RCs

- They are prenominal.
- The gap appears before the fillers.
- There are two kinds of embeddings—
  - Serial embeddings
  - Nested embeddings

31

Some facts about Mandarin RCs

- Sentences with single RCs:

….. [GAP] …….[FILLER]….

32

Some facts about Mandarin RCs

- Sentences with double RCs:

….. [GAP] ……[FILLER]….

….. [GAP] ….

Serial dependency

….. [GAP] ……[FILLER]….

Nested dependency

….. [GAP] ……[FILLER]….

33

Points of interest to us

- Comprehension of sentences with single versus double RCs
- Comprehension of double RCs with serial versus nested dependencies

34

The tasks:

- Self-paced reading
- Comprehension questions

Exp1: self-paced reading of sentences with RCs

35

Materials: Nested versus serial dependencies in head-final RCs (Mandarin)

- 迷戀|絆倒|新郎|的|女生|的|軍人|背叛了|朋友。
- 迷戀|新郎|絆倒|的|女生|的|軍人|背叛了|朋友。
- 絆倒|新郎|的|女生|迷戀|的|軍人|背叛了|朋友。
- 新郎|絆倒|的|女生|迷戀|的|軍人|背叛了|朋友。

36

Nested versus serial dependencies in head-final RCs (Mandarin)

- 迷戀 | 絆倒 | 新郎 | 的 | 女生 | 的 | 軍人 | 背叛了 | 朋友。

Nested dependency

- 迷戀 | 新郎 | 絆倒 | 的 | 女生 | 的 | 軍人 | 背叛了 | 朋友。
- 絆倒 | 新郎 | 的 | 女生 | 迷戀 | 的 | 軍人 | 背叛了 | 朋友。

Serial dependency

- 新郎 | 絆倒 | 的 | 女生 | 迷戀 | 的 | 軍人 | 背叛了 | 朋友。

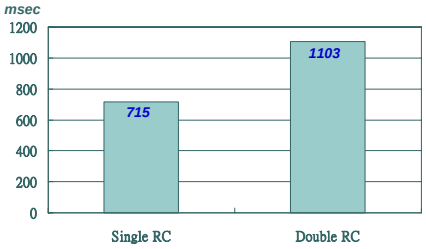
37

a. Single versus double RCs

- Average RTs
- Accuracy of comprehension

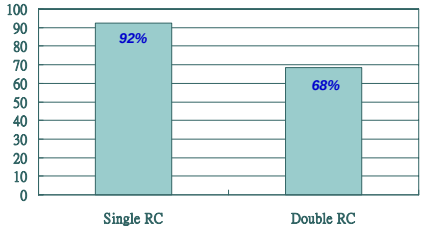
38

a. Single versus double RCs—average RTs per region



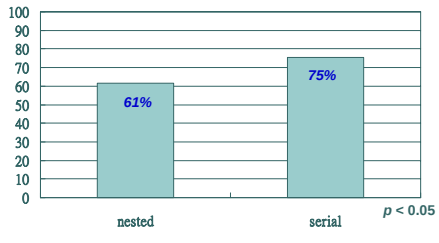
39

a. Single versus double RCs—average comprehension accuracy



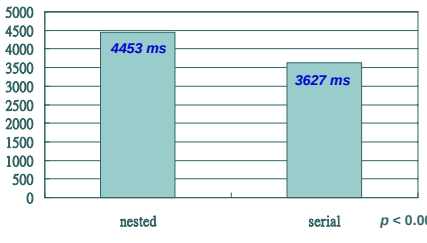
40

b. Serial versus nested dependencies—comprehension accuracy



41

b. Serial versus nested dependencies—RTs taken to answer the comprehension questions



42

summary

- Sentences that burden the working memory take longer to parse.
- These sentences are also harder to comprehend.

43

2. Complex thematic assignment in Mandarin verbal compounds

Certain linguistic constraints can be grounded on parsing.

44

2. Complex thematic assignment in Mandarin verbal compounds

The parsability of a sentence can be a confound in its grammaticality judgments.

45

2. Complex thematic assignment in Mandarin verbal compounds

- **Resultative Verb Compounds** (RVCs):

追累 *chase-tired*

(Li, 1990, 1995, 1999, *NLLT*)

我**追累**了那隻狐狸
I chase-tired that fox

Inspired by pilot studies of Yu-Bin Chen and Ray Huang.

46

我**追累**了那隻狐狸
I chase-tired that fox

(i) I chased the fox; I got tired.
(ii) I chased the fox; the fox got tired.
(iii) The fox chased me; the fox got tired.
(iv) *The fox chased me; I got tired.

47

我**追累**了那隻狐狸
I chase-tired that fox

(i) I chased the fox; I got tired.
(ii) I chased the fox; the fox got tired.
(iii) The fox chased me; the fox got tired.
(iv) *The fox chased me; I got tired.

Question for a syntactician--
Why is (iv) ungrammatical?

Questions for an experimental syntactician—
1. Why is it difficult to attain (iii)?
2. Is (iv) really ungrammatical?

(i) I chased the fox; I got tired.
(ii) I chased the fox; the fox got tired.
(iii) The fox chased me; the fox got tired.
(iv) *The fox chased me; I got tired.

causer	causer
<ag-th, pt>	<ag-th, **pt**>
<**ag**, pt-th>	<ag, **pt-th**>
<**pt**, ag-th>	<pt, **ag-th**>
***<pt-th, ag>**	**?**<pt-th, **ag**>

49

The syntactic approach

- Cheng and Huang (1994)—movement based account (structure + aspectuality)
- Li (1995, 1999)—lexicalist account
  - Thematic hierarchy
  - Causative hierarchy
- Her (2006)—lexicalist account (LFG)

50

Li (1995, 1999) cf. Her (2005)

- **Thematic Hierarchy**
- **Causative Hierarchy**
- **Well-formedness Condition** on Mapping Argument Structure to Syntax:
  - Theta roles can be assigned contrary to the thematic hierarchy if the arguments receiving them are assigned c-roles in ways compatible with the causative hierarchy. (Li, 1995:269)

51

Crucial sentences in Li (1995; cited from 1999)

(12)a. 我把那隻狐狸**追累**了

I BA that fox chase-tired-asp

'That fox is tired from my chasing it.'

zhui-lei ⟨θ1 ⟨θ2 – θa⟩⟩ → I chase that fox, the fox becomes tired.
I *that fox*
Cause Affectee

52

Crucial sentences in Li (1995; cited from 1999)

(12)b. 我今天已經**追累**了那隻狐狸了，實在走不動了。

I today already chase-tired-asp that fox LE; indeed can't walk asp

'I am already tired from chasing that fox today. I really can't walk anymore.'

zhui-lei ⟨θ1 – θa ⟨θ2⟩⟩
I *that fox* → I chase that fox, I become tired.

Crucial sentences in Li (1995; cited from 1999)

(12)c.那隻狐狸把我都**追累**了，可我還是追不上牠。

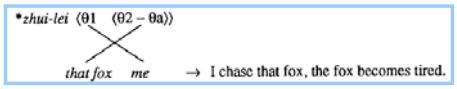
that fox BA me even chase-tired-asp but I still can't catch-up it

'I am already tired from chasing that fox, but I still can't catch up.'

zhui-lei ⟨θ1 – θa ⟨θ2⟩⟩
that fox *me* → I chase that fox, I become tired.
Cause Affectee

54

- (12)d.



55

The processing approach

HYPOTHESES:

- Perceptual factor—NVN sequencing and thematic/causative mapping (Bever, 1970; Townsend & Bever, 2001)
- Multiply embedded null causatives are hard to parse

56

NVN sequences (Bever, 1970)

- Tend to be taken as **Subject-Verb-Object**
- Then interpreted as **Agent-Action-Patient**

57

Bever (1970), Townsend and Bever (2001)

main-clause analysis

N **V** **N**

The horse raced past the barn fell.

Theme

⬇ ***reduced RC analysis***

Patient

58

Chinese RCs with object topicalization (Lin, 2006)

RC					HEAD	
e 放棄過	好幾位	男友	的		女演員	從不後悔
V1	CL	N1	DE		**N2**	V2

forsake ASP many CL boyfriend DE actress always not regret

'The actress who has forsaken many boyfriends never felt regretful.'

59

The topicalized object is animate.

RC					HEAD	
e 放棄過	好幾位	男友	的		女演員	從不後悔
V1	CL	N1	DE		**N2**	V2

forsake ASP many CL boyfriend DE actress always not regret

'The actress who has forsaken many boyfriends never felt regretful.'

N **V** **N**

RC					HEAD	
男友	e 放棄過	好幾位	t	的	女演員	從不後悔
N1	**V1**	CL	t	DE	**N2**	V2

boyfriend forsake ASP many CL t DE actress always not regret

'The actress who has forsaken many boyfriends never felt regretful.'

60

The topicalized object is *in*animate. (Lin, 2006)

RC				HEAD	
e 放棄過	好幾次	機會	的	女演員	從不後悔
V1	CL	N1	DE	N2	V2

forsake ASP many CL chance DE actress always not regret

'The actress who has given up many chances never felt regretful.'

N V N

RC					HEAD	
機會	e 放棄過	好幾次	t	的	女演員	從不後悔
N1	V1	CL	t	DE	N2	V2

chance forsake ASP many CL t DE actress always not regret

'The actress who has given up many chances never felt regretful.'

61

我追累了那隻狐狸
I chase-tired that fox

(i) I chased the fox; I got tired.
(ii) I chased the fox; the fox got tired.
(iii) The fox chased me; the fox got tired.
(iv) *The fox chased me; I got tired.

Question for a syntactician--
Why is (iv) ungrammatical?

Questions for an experimental syntactician—
1. Why is it difficult to attain (iii)?
2. Is (iv) really ungrammatical?

The processing approach

HYPOTHESES	PREDICTIONS
Perceptual factor—NVN sequencing	"Agent Verb Patient" is preferred.
Perceptual factor—NVN sequencing	An untypical agent enforces a causative parse.
(Multiply embedded) null causatives are hard to parse	Causatives are hard.
(Multiply embedded) null causatives are hard to parse	Multiply embedded null causatives are beyond WM.

63

Some preliminary evidence

64

The processing approach

HYPOTHESES	PREDICTIONS
Perceptual factor—NVN sequencing	"Agent Verb Patient" is preferred.
Perceptual factor—NVN sequencing	An untypical agent enforces a causative parse.
(Multiply embedded) null causatives are hard to parse	Causatives are hard.
(Multiply embedded) null causatives are hard to parse	Multiply embedded null causatives are beyond WM.

65

Huang's (2007) survey

Ratings and interpretations from 10 native speakers of Chinese (in a questionnaire) with sentences like: 桃桃追累幼幼了(Taotao chase-tired Youyou),

- **None** reported causatives as the primary meanings. That is, N1 is treated as the agent-theme, N2, the patient.

66

The processing approach

HYPOTHESES	PREDICTIONS
Perceptual factor—NVN sequencing	"Agent Verb Patient" is preferred.
Perceptual factor—NVN sequencing	An untypical agent enforces a causative parse.
(Multiply embedded) null causatives are hard to parse	Causatives are hard.
(Multiply embedded) null causatives are hard to parse	Multiply embedded null causatives are beyond WM.

67

Huang's (2007) survey

Ratings and interpretations from 10 native speakers of Chinese (in a questionnaire) with sentences like:
這件事忙累幼幼了(This task busy-tired Youyou),
媽媽的嘮叨煩死小寶了(Mom's nagging annoy-dead Xiaobao)

- **All** reported causatives as the primary meanings.

68

The processing approach

HYPOTHESES	PREDICTIONS
Perceptual factor—NVN sequencing	"Agent Verb Patient" is preferred.
Perceptual factor—NVN sequencing	An untypical agent enforces a causative parse.
(Multiply embedded) null causatives are hard to parse	Causatives are hard.
(Multiply embedded) null causatives are hard to parse	Multiply embedded null causatives are beyond WM.

69

我追累了那隻狐狸
I chase-tired that fox

NVN
Agent-V-Patient
+ Causative

(i) I chased the fox; I got tired.
(ii) I chased the fox; the fox got tired.
(iii) The fox chased me; the fox got tired.
(iv) *The fox chased me; I got tired.

Structural + Processing Account

70

我追累了那隻狐狸
I chase-tired that fox

(i) chased(I, fox)&tired(I)
(ii) v_{cause}(chased(I,fox), tired(fox))
(iii) v_{cause}(I, chase(fox,I)&tired(fox))
(iv) v_{cause}(I, v_{cause}(chased(fox,I), tired(I))

71

我追累了那隻狐狸
I chase-tired that fox

(i) I chase-tired fox
(ii) I chase-tired fox
(iii) I CAUSE fox chase-tired I
(iv) I CAUSE fox chase-tired I

72

我追累了那隻狐狸
I chase-tired that fox

(i) chased(I, fox)&tired(I)
(ii) v_{cause}(chased(I,fox), tired(fox))
(iii) v_{cause}(I, chase(fox,I)&tired(fox))
(iv) v_{cause}(I, v_{cause}(chased(fox,I), tired(I))

+ Causative

*
ConfoundI

The processing approach

HYPOTHESES	PREDICTIONS
Perceptual factor—NVN sequencing	"Agent Verb Patient" is preferred.
Perceptual factor—NVN sequencing	An untypical agent enforces a causative parse.
(Multiply embedded) null causatives are hard to parse	Causatives are hard.
(Multiply embedded) null causatives are hard to parse	Multiply embedded null causatives are beyond WM.

我追累了那隻狐狸
I chase-tired that fox

我追累了那隻狐狸 I chase-tired that fox	NV1N: Ag V PT	*v_{cause} Ø	NV2: Th V	
I chased the fox; I got tired.				1
I chased the fox; the fox got tired.		*	*	2
The fox chased me; the fox got tired.	**	*	*	3
*The fox chased me; I got tired.	**	**		4

implications

- When a sentence/word is beyond comprehension, it is difficult to test its grammaticality.
- Grammaticality judgments from language users whose WMs are limited (linguists included) are not necessarily reliable.

Other similar approaches

- "Superiority effects are the result of interacting processing preferences." (Arnon et al., 2005; Frazier & Clifton, 2005)
- Processing evidence for strong versus weak islands (Kluender, 1998)
- Set theoretical approach to coordination (Cowart, yesterday)

Concluding remarks

- These data rarely/never appear in the corpora given that they are not easy to parse/produce.
- Cases that we take as grammatically ill-formed may actually be "difficult" sentences.

Concluding remarks

- Sentences that do not make sense to the "miniJudges" are likely to be taken as ungrammatical.

79

Concluding remarks

- Only when processing/parsing accounts do not work should we resort to syntactic analyses!

80

Thank you.

81