Discourse Analysis

Some Preliminaries

- Broadly speaking, studies on human language are divided into **two major schools**.

- One school of studies investigates a language in close connection with the social and cultural environments in which the language is learned or employed.

- The other school of studies asserts that the major goal of linguistics is to define and explain the linguistic competence of an imaginary “ ideal speaker” to produce and understand language.
- Linguists working within this paradigm also assert that the core of competence is innately determined and is to be defined independently of any social or contextual factors.

Context

- I've got a flat tyre.

In a garage, this might be taken to mean that I need help; or addressed to a friend with a car, that I need a lift; or as a response to a request for a lift from a friend without a car, that I cannot give him/her a lift.

- “ 你真坏 !”

- Types of context:

linguistic context/co-text
situational/physical context
social/cultural context

- Linguistic context

It refers to specific parts of an utterance (or text) near or adjacent to a unit which is the focus of attention.

e.g. bank

1. used in a sentence together with words like *steep* or *overgrown*
2. hear somebody say that she has to *get to the bank to cash a check*

- This exchange took place between two young Englishmen on a train. Speaker A was reading a newspaper.

A: What’s this lump they’re always on about?

B: Read it out.

A: [reads aloud from paper] *Inland Revenue crack down on lump.*

B: Oh, isn’t it something to do with casual labour on building sites? The way they’re paid--- tax evasion and that?

- Notice that speaker B did not list all the possible senses of lump(a shapeless mound; a hard swelling; a heavy, dull person;). Instead, he asked for sufficient context to enable him to tell his friend what, in his opinion, the writer of this particular article meant by the word lump.

- Situational Context/physical context

e.g. bank

If you see the word BANK on the wall of a building in a city, the "physical" location will influence your interpretation.

A: Got it's hot in here.

B: Is it?

C: Yeah. Really. Are you shutting out this lovely sunshine?

B: It's getting in my eyes.

A: Oh no!

D: Yeah. Not used to that are you?

B: No. What's that? Psycholinguistics?

Systemic Functional Approach

- Michael Halliday and Others

- Theoretical claims about language
    1. functional
    2.semantic
    3. contextual
    4. semiotic

- Language use is functional;
- Its function is to make meanings;
- These meanings are influenced by the social and cultural context in which they are exchanged;
- The process of using language is a semiotic process, a process of making meanings by choosing.

Semiotic System

- **Semiotic System of Traffic Lights**
- stop → red
-
- slow down → amber
-
- go → green

(meaning) (realization)

Language as a semiotic system

system of lexical choice

"When I got home from work yesterday, I could not , believe what my (progeny) had done!"

- kid, child ,brat, darling, son, boy…

Specify sex *son, boy*

Don't specify sex *child, brat, darling*

System of grammatical choice

- Number — singular / plural
- Person — first / second / third

What is discourse?

discourse

talk (spoken discourse)

text (written discourse)

discourse analysis

text linguistics

Discourse studies are about *talk and text in context.*

What is discourse analysis? (DA)

- **Michael Stubbs(1983)**

In his book, discourse analysis is defined as:

1) concerned with language use beyond the boundaries of a sentence/utterance,

2) concerned with the interrelationships between language and society and

3)concerned with the interactive or dialogic properties of everyday communication.

What is discourse analysis? (DA)

Teun A. van Dijk (1997)

Three dimensions of discourse analysis

1. language use
2. the communication of beliefs (cognition)
3. interaction in social situations

Types of discourse analysis

1) those which focus on ***discourse*** "itself", that is , on structures of text or talk;
2) those which study discourse and communication as ***cognition***
3) those which focus on ***social structure and culture***

This triangle of ***discourse—cognition—society*** is indeed the site of multidisciplinary discourse analysis

Cohesion

e.g.

I bought a Ford. A car in which President Wilson rode down the street was black. Black English has been widely discussed. The discussions between the presidents ended last week. A week has seven days. Every day I feed my cat. Cats have four legs. The cat is on the mat. Mat has three letters.

Does this sequence represent a complete, unified linguistic event?

Does each clause in the sequence contribute something to the text being a meaningful whole?

Do the clauses in the sequence hang together to make them a unified whole?

When we say the clauses don’t hang together, we are reacting to two dimensions of the sequence: ***its coherence and cohesion.***

- Cohesion refers to these surface-structure features of a text which link different parts of sentences or larger units of discourse.

e.g.

The man went to town. ***However****, he did not stay long.*

- Coherence refers to the main principle of organization postulated to account for the underlying functional connectedness or identity of a text. It involves the study of such factors as the language user's knowledge of the world, the inferences they make, and the assumptions they hold.

  e.g. A: there's the doorbell.

  B: I'm in the bath.

  A: Ok. I'll open it.

- **Different types of cohesion**
- **Grammatical cohesion and lexical cohesion**
- **Grammatical cohesion:** reference,ellipsis, substitution and conjuntion

- Reference

  It refers to how the writer/speaker introduces participants and then keeps track of them once they are in the text.
- Participants are the people, places and things that get talked about in the text.
- *e.g.* ***John*** *was generally not a hater.* ***He*** *spoke tolerantly of his foes.*

- Types of reference

exophora

endophora

anaphora

cataphora

- exophora: pointing something to the world
  *e.g. Look at **that**.*
- endophora: pointing inwards to the text

- anaphora: pointing backwards in the text.

- cataphora: pointing forwards in the text.

*e.g. Wash and core six cooking apples. Put **them** into a fireproof dish. (anaphora)*
*The point of my story is **this**: we should donate blood. (cataphora)*

Reference items:

- 1. the definite article: the

e.g. Bungling ram-raiders tried to smash their way into a furniture shop—using a stolen Mini.But **the** tiny motor just bounced off **the** store's plate-glass window.

- 2. demonstrative pronouns: that, this, these, those

  *e.g. I'll put you through into the looking-glass house. How could you like **that**?*

- 3. third-person pronoun: he, she, they, it

  *e.g. Cholera first struck England in 1832. **It** came from the East.*

- 4. comparative reference: *more, another, other,different, else, the same, similarly,etc.*
- *e.g.*

*John's daughter had premature jaundice and things. Her **other** problems included….*

*Something **else** that was happening to her was….*

*A **different** but equally common problem is….*

- 5. locational reference: there, then, here, now, etc.

  It involves not the identification of a participant in a text, but the identification of a location in time or space. Locational referents to nearby time or space are frequently retrieved exophorically, while locational items referring to distant time or space are often endophorically retrieved.

  e.g. I went to the clinic. ***There*** *I found my daughter. (retrieved anaphorcially to the clinic)*

- 6. bridging reference

- definition

A bridging reference item is one that is used to establish a link of association with some preceding expression in the same text via the addition of background assumptions.

e.g. John walked into a music hall. ***The piano*** *was made in the nineteenth century.*

- properties
- ① The anaphor, which is usually a definite NP, must occur in the appropriate context of its “antecedent”, which is usually an indefinite NP.

- ② There is some semantic and/or pragmatic relation between the anaphor and its “antecedent.”

③ The anaphor and its “antecedent” do not stand in a strictly coreferential relation. Rather they are linked to each other via the addition of pragmatic inference of some kind.

- Ellipsis

Another form of anaphoric cohesion in the text is achieved by ellipsis, where we presuppose something by means of what is left out.

e.g. —Oh, you didn't have to make a special trip for that, Mum.

— I didn't.

- Types of ellipsis in English

① nominal ellipsis;

②verbal ellipsis ;

③clausal ellipsis;

- nominal ellipsis

e.g.

She wore the blue dress, but the pink suits her better.

—Have some wine.

— There isn't any.

- verbal ellipsis

  e.g.

  Alison: Do you understand?

  Helen: Do you?

  Alison: I've tried to.

- clausal ellipsis
- ①yes/no ellipsis
- ②wh-ellipsis

yes/no ellipsis:

a, the whole clause

b, part of the clause

e.g.

— *Can you row?*

— *Yes.(I can row.)*

— *Is that all?*

— *No. (That's not all.)*

---Must a name mean something?

---Of course it must. (mean something)

---Hold your tongue!

---I won't! (hold my tongue)

- wh-ellipsis:
- a. the whole clause
- b. part of the clause.

e.g. ---I think you ought to tell me who you are first.

--- Why?(ought I to tell you who I am)

---They are at it again.

---Who? (who are at it again?)

---The lion and the unicorn, of course.

--- They are at it again.

--- Who are? (at it again)

--- Who can untie this knot?

--- I can.(untie this knot)

- **Substitution**

Sometimes an explicit indication may be given that something is omitted, by the use of a substitute form.

e.g. --- I've lost my watch.

*--- Get a new **one**.*

- Types of substitution in English

①nominal substitution;
②verbal substitution;
③clausal substitution;

- **nominal** substitution

The most common nominal substitute is <u>one</u>, plural <u>ones</u>.

e.g.

--- That's a joke. I wish you had made it.

---Why do you wish I had made it? It's a very bad ***one****. (a very bad joke)*

"And you'll meet lots of nice men there, gentlemen, not like the ***ones****(=men) I knew."*

- The negative substitute form in these cases is “none”.

  e.g.

  --- “What wonderful things have you told her about me?”

  *--- **None**.(no wonderful things)*

- Besides one(s), same, that , those and possessive pronouns also act as substitute forms for nominal groups.

e.g. --- Can I have a cup of black

coffee with sugar, please?"

--- Give me the ***same****, please.*

American food is not the same as the English ***kind****.(kind=food)*

Between the hounds and the horses and the twins there was a kinship deeper than ***that*** *of their constant companionship.*

---- Take off your hat.

---- It isn't ***mine****.(my hat)*

---- Stolen!

---- I keep them to sell. I've ***none*** *(no hats) of my own.*

- **verbal** substitution

  Verbal substitution is by means of the verb **do**, which can substitute for any verb provided it is active not passive.

  e.g.

  ---- Does it hurt?

  *---- Not any more. It was **doing** last night.*

---- Do they buy their drinks at the local supermarket?

*---- No ,but we **do**.*

---- Have the children gone to sleep?

*---- I think they must have **done**.*

- **clausal** substitution

The common clause substitutes are **so** and **not.**

In certain contexts these substitute forms are used:

① following if — if so, if not.

②as a reported clause: he said so, he said not.

③in the context of modality: perhaps so, perhaps not.

e.g.

He will not die unless we abandon him, if ***so*** *(if we abandon him), we are indeed answerable for his blood.*

If I like being that person, I'll come up, if ***not****(=if I don't like being that person), I'll stay down here till I'm somebody else.*

--- Are the Browns coming to dinner?

--- I hope ***so****.*

---I dare say you never even spoke to Tim!

---Perhaps ***not****.*

--- The Yankees will win the game.

--- All my friends say ***so****.*

- With a declarative response, there is a change of subject only, we may have substitute **so**, **nor** in initial position.
- e.g.

  A: *Of course, you know your ABC?*

  B: To be sure I do.

  A: ***So*** *do I.*

  A: I haven't the slightest idea.

  B: ***Nor*** *have I.*

If the Subject is unchanged, so that the focus is on the Finite, the order is S ^ F.

e.g.

--- I want to be a Queen.

*--- **So** you will. (be a queen.)*

--- He has arrived.

*--- **So** I have heard. (I have heard that he has arrived.)*

- **Conjunction**
- The cohesive pattern of conjunction refers to how the writer creates and expresses logical relationships between the parts of a text.

- **Types of conjunctive relations: elaboration, extension, and enhancement**

- 1. Elaboration is a relationship of restatement or clarification, whereby one sentence is a re-saying of a previous sentence.
- Common conjunctions used to express this relation include: in other words, that is (to say), I mean (to say), for example, for instance, thus, to illustrate, to be more precise, actually, in fact, as a matter of fact.

- *e.g.*

  John's daughter was very sick. ***In fact****, she needed a blood transfusion.*

  My daughter had jaundice. ***I mean*** *she was as yellow as butter.*

- 2. Extension is a relationship of either addition (one sentence adds to the meanings made in another), or variation (one sentence changes the meanings of another, by contrast or by qualification).

- Typical conjunctions include: and, also, moreover, in addition, nor, but, yet, on the contrary, instead, apart from that, except for that, alternatively.

e.g.

He has a daughter. ***And*** *he has got 3 sons.*

We could give blood tomorrow. ***Or*** *maybe the day after would be better.*

3. Enhancement refers to ways by which one sentence can extend on the meanings of another, in terms of dimensions such as time, comparison, cause, condition or concession.

- Common temporal conjunctions include: then, next, afterwards ,just then, at the same time, before that, soon, after a while, meanwhile, until then, up to that point.
- *e.g.Diana spent a month in Geneva.* ***Then*** *she left Europe for Africa.*

- Comparative conjunctions include: likewise, similarly, in a different way.

*e.g. Diana had her blood tested. **Likewise**, the Swiss donors had to be checked too.*

- Causal conjunctions include: so, therefore, consequently, hence, because of that, for, in consequence, as a result, on account of this, for that reason, for that purpose.

*e.g. Diana gave blood. **As a result,** her daughter recovered.*

- *Conditional conjunctions include: in that case, otherwise, under the circumstances, if not.*

  e.g. Diana gave blood. ***Otherwise*** *her daughter wouldn't have recovered.*

- Concessive relations are expressed by: though, still, even so. however, nevertheless, all the same.

  e.g. ***Though*** *Diana doesn't like giving blood, she does it all the same.*

- **Lexical cohesion**
- It refers to how the writer/speaker uses lexical items(nouns, verbs, adjectives, adverbs), and event sequences (chains of clauses and sentences) to relate the text consistently to its area of focus. Lexical relations analysis is a way of systematically describing how words in a text relate to each other, how they cluster to build up **lexical sets or lexical strings**.

- Lexical relations operate between units which encode lexical content, i.e. the open class items of **nouns, main verbs, adverbs and adjectives**. Grammatical words, or closed-class items, such as prepositions, pronouns, articles, auxiliary verbs, etc. do not encode lexical content and cannot enter into lexical relations.

- There are two main kinds of lexical relations that we can recognize between words:
- **1. taxonomic lexical relations**: where one lexical item relates to another through either class/sub-class(rodent—mouse) or part/whole (tail—mouse) relations. Although most frequently these relations link lexical items which refer to people, places, things, and qualities, and so are expressed in nominal groups, taxonomic relations can also link processes (verbs) (eat—nibble)

- **2. expectancy relations: w**here there is a predictable relation between a process (verb) and either the doer of that process, or the one effected by it (done-to). These relations link nominal elements with verbal elements.

  (mouse-squeak, nibble-cheese).

- **1.Taxonomic lexical relations**

Words which are taxonomically related may be related through either classification or composition.

1) Through classification: this is the relationship between a superordinate term and its members, or hyponyms. Classification is the “X”is a type of “y”relationship .The major kinds of classification relations are:

- a, co-hyponymy: when two(or more) lexical items used in a text are both members of a superordinate class.

  e.g. ***jaundice, pneumonia*** *(both items are members of the superordinate class, illness.)*

- b, class/sub-class: when two (or more) lexical items used in a text are related through sub-classification:

e.g. ***illness, headache*** *(here the relationship is superordinate term to hyponym)*

- c. contrast: when two (or more) lexical items encode a contrast relationship.

e.g. ***wet ,dry*** *(the antonymy relationship)*

- d. similarity: when two(or more) lexical items express similar meanings.

  There are two main sub-types:

  i)synonymy: when two(or more) lexical items essentially restate one another.

  e.g. ***clinic, hospital***

  ii) repetition: when a lexical item is repeated.

  e.g. ***blood, blood***

*2) Through **composition**: Composition is the part/whole relationship between lexical items which are meronyms or co-meronyms. There are two possible types:*

- Meronymy: when two or more lexical items are related as whole to part (or vice versa)

  e.g. ***body: neck, head, arm, leg.***

- Co-meronymy: when two (or more) l*exical items are related by both being parts of a common whole.*

  e.g. ***trunk, branch, crotch, limb, leaf***

- **2. expectancy relations**

  The relation may operate between an action and the typical (expected) doer of that action.

  doctor/operate

  baby/cry

  dog/bark

  rat/squeak

- or the relation may operate between an action/process and the typical(expected) participant effected by that action:

  smack/child

  eat/dinner

  play/piano

- The predictability relationship between an event/process and the typical location in which it takes place may also be described as an expectancy relation:

  transfusion/hospital

  classroom/school

- Expectancy can also be used to capture the relationship between the individual lexical items and the composite, predictable, nominal group they form:

  blood/donor

  blood/transfusions

- **The interdependence of grammatical and lexical cohesion**

In a text of non-minimal size, there normally occur many threads of semantic relations based on grammatical and lexical cohesive devices. However, to be effective, grammatical cohesion requires the support of lexical cohesion, and the lexical cohesion, in its turn, depends on grammatical cohesion.

- e.g.

(1) John gets up early. We bought him a tie. He loves peaches. My house is next to his.

(2) A cat is sitting on a fence. A fence is often made of wood. Carpenters work with wood. Wood planks can be bought from a lumber store.

In (1), we have **grammatical cohesion not supported by lexical cohesion.**

In (2), the reverse is operating. If we have to accept them as texts, they are odd rather than typical.

Exercises:

Text 1 (Analysis for reference)

Alice looked on with great interest as the King took an enormous memorandum book out of his pocket, and began writing. A sudden thought struck her and she took hold of the end of the pencil, which came some way over his shoulder, and began writing for him.

The poor King looked puzzled and unhappy, and struggled with the pencil for some time without saying anything; but Alice was too strong for him, and at last he panted out, “My dear! I really must get a thinner pencil. I can’t manage this one a bit; it writes all manner of things that I don’t intend --”

“What manner of things?” said the Queen, looking over the book (in which Alice had put “The White Knight is sliding down the poker. He balances very badly”). “That’s not a memorandum of your feelings!”

Text 2 (Analysis for ellipsis and substitution)

"Being so many different sizes in a day is very confusing."

"It isn't."

"Perhaps you haven't found it so yet; but when you have to turn into a chrysalis—you will do some day, you know—and then after that into a butterfly, I should think you'll feel it a little queer, won't you?"

"Not a bit."

"Well, it would feel very queer to me."

"You! Who are you?"

"I hardly know, sir, just at present."

"So you think you're changed, do you?"

"I'm afraid I am, sir."

Discourse Anaphora

- (1) Xiaoming^{x} gei Xiaohua^{2} ***ziji*** de hua.
  (intrasentential anaphora)
- （2） John said that he would come.
  (intrasentential anaphora)
- (3) John came in. He sat down.
  (discourse anaphora)
- （4） John^{x} and Bill^{2} came in. John^{x} said that he would come.
  (discourse anaphora)

Discourse Anaphora in Chinese Conversation

- Focusing on discourse anaphora in its prototypical use: how the establishment, shift and maintenance of reference to a third-person singular human entity is done in conversation

1.1 Anaphoric distribution in conversation

- the basic distributional pattern of anaphora for the maintenance of reference in Chinese conversation

  NP^{x}…

  P^{x} …

  Ø^{x}…

1. A: Ni jiali xianzai
2. Xiao Tian xianzai hai hao a?
3. B: Xiao Tian hai keyi
4. A: Ta xianzai haishi zai lao danwei
5. gongzuo a?
6. B: Ø hai zai nabian

- The above pattern has two variants:
- NP^{x}…

  $Ø^{x}$…

- NP^{x}…

  P^{x} …

1. A: Shui dao Hangzhou lai guo de?
2. B: En, Dahua jinnian lai guo de
3. A: Ao, Ø shi hui jia tanqin
4. Haishi
5. B: Ø jiushi hui jia tanqin

1. A: Liu Laoshi chu le ben shu
2 B: Ta shi ting yonggong de.

- The basic distributional pattern of anaphora for the shift of reference in Chinese conversation

NP^{x}…

NP^{2}…

1. A: Cai Lin zai yinhang gongzuo
2. B: Ta airen ne?
3. A: Ø haoxiang zai xintuo gongsi
4. Ao Cai Lin qunian hai qu le tang
5. Xianggang.

- 1. A: Jiali ma ye yong le yi ge yongren
- 2. Yongren ne ye hui zuo cai xi yishang
- 3. Ø nengli bijiao,bijiao shi gao de
- 4. Ø guan xiaohair

From the above discussion, we can derive the basic distributional pattern of anaphora in Chinese as follows:

- 1. Establishment of reference tends to be achieved through the use of a lexical NP.
- 2. Shift of reference tends to be achieved through the use of a lexical NP.
- 3. Maintenance of reference tends to be achieved through the use first of a pronoun and then of a zero anaphor.

1.2 Anaphoric production in conversation

- Addressing the question of anaphoric production, namely what contributes to the speaker's choice of a particular anaphoric form at a particular point in Chinese conversation
- There is general consensus in the literature that there are two principles operating in determining discourse anaphora: a principle of clarity and a principle of economy.

- The Clarity (C-)principle

  “say as much as you must to avoid ambiguity”

  i.e. use a full NP whenever you have to.

- The Economy (E-)principle

  “say as little as you can get away with (given C)”

  i.e. use a PRO-form whenever you can.

- 1.2.1 Establishment of reference

When a speaker intends to introduce a person into conversation, he is faced with a problem: for any person to whom reference is to be made, there is a (potentially infinite) set of possible referential expressions each of which , by a correspondence test, is “correct” or “appropriate” and therefore could be used to identify that person.

On any actual occasion of use, however, it is not the case that just any member of that set is “right” or “appropriate”. Therefore, the speaker has to select an “appropriate” form from that set, but what constrains the choice?

- According to Sacks & Schegloff(1979), all referential work in conversation is essentially “recipient-designed” . This indicates that the speaker, in deciding on a particular referential form, has to ensure that it is one that can serve for the hearer to “locate” the intended referent. Thus, anaphoric production in conversation depends crucially on the assumptions made by the speaker about how the hearer will recognize the intended referent.

- To begin with, the speaker has to decide whether the person to be introduced into conversation is known to the hearer. If the person is judged to be known to the hearer, a definite expression is used; otherwise, an indefinite expression is employed.

- 1. A: Sanmao ba xianzai?

  2. B: Sanmao zai Shanghai Jiguang

  3. Yanjiusuo

- 1. A: Ø zhao le ge ayi

  2. Ayi ye zou le

- Next, in cases where the first mention of a person is encoded in definite expressions, the speaker has to decide whether the referent is to be identified through a definite description or a proper name. Generally speaking, a definite description will be used when the referent is to be identified via the relevant role, otherwise a proper name is used.

1. A: Ni baba shenti ye keyi
2. Ø hai dao-dao HUangshan wan
3. le yi xia

Proper names are the predominant referential form for introducing known human referents into Chinese conversation.

Firstly, a proper name is often chosen over a definite description even in cases where the role of the referents is clearly relevant and thereby a definite description might have usefully be used.

- (Yaxiang is the president of the student union)

1. A: Zhe shi a Ø zuihao haishi wenwen
2. Yaxiang

Secondly, a proper name is often used immediately following a definite description.

1. A: Name zhixifu ne Xiufang ye tebie hao

Thirdly, in cases where the speaker has a "word-search" problem, and cannot find a referential form for the person he has in mind, it is normally the proper name that is solicited.

1. A: Ta ne neige Liu shenme?
2. B: Liu Yihua
3. A: Liu Yihua ai
4. B: Liu Yihua zai nar gan de ting hao de
5. A: Ao, ao

Some interesting cases

- 1) The speaker is uncertain about the adequacy of the referential form being used for achieving recognition, and so inserts a recognition check sequence immediately following the use of (usually) a proper name.

1. A: Jisuanjixi you yi ge ZhangMinghui
2. bu zhidao ni renshi Ø a?
3. B: Ai wo renshi Ø a
4. A: Ai Zhang Minghui meiyou ti shang
5. Ø ye you yijian
6. B: Danshi Zhang Minghui hai bei
7. ping shang xianjin jiaoshi a
8. Shenmede
9. Dui bu dui

1. A: Yu Xiaobin ni zhidao bu zhidao?
2. B: (pause)
3. A: Ta shi gang lai de
4. B: Shi ma?
5. C: Ø xue sheme de?
6. A: Ø xue guoji jingjifa de
7. C: Ø ye shi Beida de?
8. A: Ø Beida Falvxi de

- 2) The speaker anticipates that the hearer is able to recognise the referent, but in fact the hearer cannot. In such cases, an other-initiated recognition search sequence often occurs.

1. A: Xiao Liu gang lai guo
2. B: Neige Xiao Liu?
3. A: Liu Jianguo
4. B: Ao

- 3) The speaker thinks that the referent is unknown to the hearer, and so uses a “non-recognitional” form; but in fact the referent is known to the hearer. In such cases, the hearer often seeks to confirm his suspicion by providing the referent’s name or by asking for it.

1. A: Shang xingqi you lai le ge nvsheng
2. B: Ø shi bu shi jiao Shang Subo?
3. A: Shia ni ye renshi Ø?
4. B: Dui wo ye renshi Ø

- 4) The speaker has “word-search” problem and cannot find the name for the referent. He may then use a description to help recover the name from his own memory or to solicit it from the hearer.

1. A: Haiyou nage fen dao Beijing qu de
2. Neige neige bu da shuohua de
3. Jiang shenmede?
4. B: Ao Jiang Yunan
5. A: Ai Jiang Yunan dui

- 5) There are cases where a proper name is employed for subsequent use even when its referent is not known to the hearer.

1. A: Ta hai zai nabian hai pengdao yi
2. ge yuanlai zai zhebian Ruishi de
3. yi ge liuxuesheng
4. Ni hai jide Ø ba?
5. B: (pause)
6. A: Ø gezi bu tai gao
7. B: Wo yinxiang yidian dou bu shen le
8. A: Ø ting yonggong de
9. B: En
10. A: Ø jiao neige Fan Ke
11. B. En

1.2.2 Shift of reference

- The shift of reference in Chinese conversation is done through the use of a lexical NP.

1.A: Ta airen ma ye shi sanbandao nvgong

2. Ø Sheng le yi ge erzi

3. Erzi dao shi man congming de

4. Hai hui changge

5. Tiaowu ye hui de

- Pronouns and zero anaphors are occasionally used for the shift of reference in Chinese conversation

1. A: Ni zai zhebian hai pengdao nage
2. Zhuang Guoou la?
3. B: Ø Meiyou a
4. Ta xianzai liu zai Wenyansuo la

1.2.3 Maintenance of reference

- The Maintenance of reference in Chinese conversation tends to receive first a pronoun and then a zero anaphor encoding.

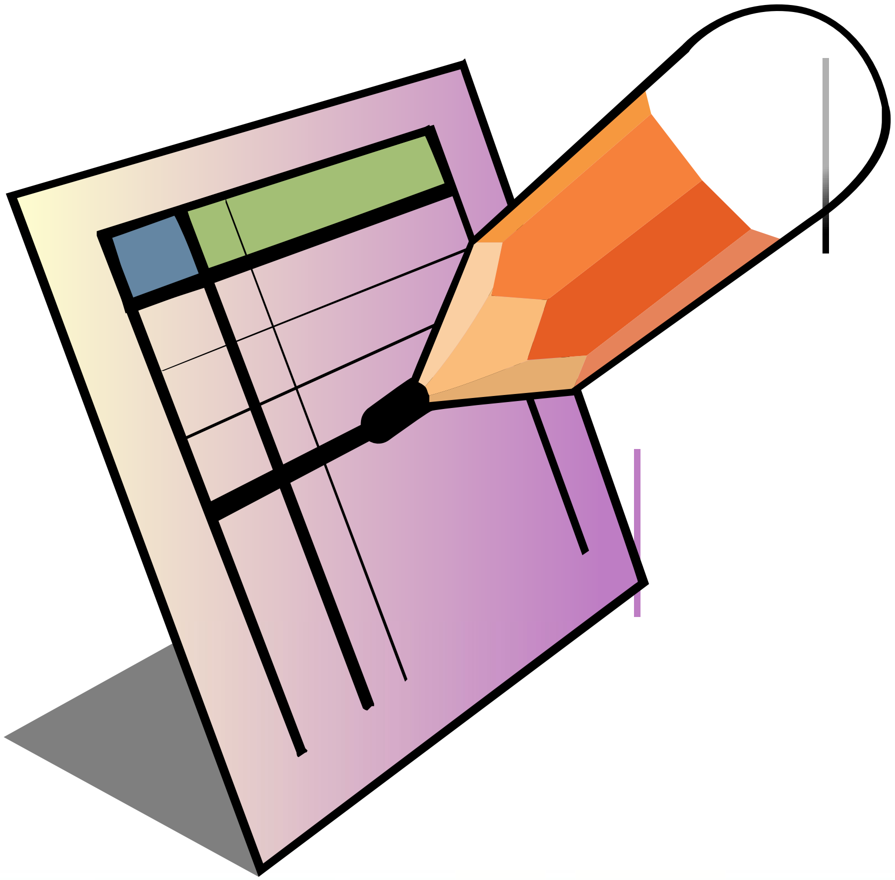
- 1. A: Li Ruijing lai guo le
  2. B: Ao
  3. A: Li Ruijing
  4. B: Ta xianzai hai zai Minhang?
  5. A Ta haoxiang ye kaoqu le
  6.   yanjiusheng le
  7. B: Dui dui dui wo ting wo ting
  8.   xiangbuqi ting shui shuo guo

1. A: Na ta nage ta- ta qizi gan
2. shenme ne?
3. Ta qizi ma zuo gongren
4. Ø suliao changli gongzuo
5. Ø ye bijiao mang

Linguistic data

- Types of linguistic data
- Linguistic data: intuitive data
  non-intuitive data
- Intuitive data: introspective data
  elicitation data
- Non-intuitive data: anecdotal data
  corpus data

Collecting and Transcribing Data



Collecting Data

- If you are going to analyze any kind of discourse, you need material ('data') to work on. Collecting data involves recording speech, gathering written texts, obtaining permission from speakers and writers to use their texts, and keeping careful records about the texts collected and the individuals from whom they were obtained.

- Compared with collecting spoken language data, most written materials are relatively straightforward to collect.
- Spoken language data for discourse analysis consist in the first instance of recordings (audio or video) of people talking; the next part of the process is to construct a transcript, a representation of talk in written form which will serve as the main input to analysis.

- Speech is the primary mode of human communication. As a result, all kinds of speech have evolved: there are not only spontaneous multi-party dialogues but scripted and unscripted monologues, radio and television interviews, telephone conversations, class lectures, and so forth.

- There are some methodological considerations that must be addressed in the collection of speech. In collecting any kind of speech, the central concern is obtaining speech that is ‘natural’. This is particularly important when gathering spontaneous multi-party dialogues, such as informal talks between two or more individuals. If multi-party dialogues are not carefully collected, the result can be a series of recordings containing unnatural speech.

- Collecting ‘natural’ multi-party dialogues involves more than simply turning on a tape recorder and having people converse. As anyone who has collected language data knows, as soon as you turn on a tape recorder and tell people to speak, you have to deal with the common scientific problem known as the ‘observer’s paradox’: by observing something, you change its natural behavior. And this is especially true with speech because many individuals are quite insecure about the way that they speak, and when they know that their speech is being monitored, they may adjust their speaking style and attempt to speak what they perceive to be ‘good’ Chinese or English or any other language.

- To avoid this problem, those creating earlier corpora recorded people surreptitiously, and only after recordings were secretly made were individuals informed that they had been recorded. While it may have been acceptable and legal back in the 1950s and 1960s to record individuals without their knowledge, now such recordings are illegal in some places and considered unethical by most linguists. It is therefore imperative that anyone recording speech not only informs individuals that they are being recorded but obtains written permission from the individuals to use their speech in the corpus.

- Since it is not possible to include surreptitious speech in a corpus, does this mean that non-surreptitiously gathered speech is not natural? This is an open question, since it is not possible to answer it with any definite certainty: so little speech has been collected and included in corpora that it is not really possible to know empirically precisely what natural speech is really like. However, there are ways to increase the probability that the speech included in a corpus will be natural.

- First, before individuals are recorded, they should be given in written form a brief description of the project they are participating in. In this description, the purpose of the project should be described, and it should be stressed that speech is being collected for descriptive linguistic research, not to determine whether those being recorded are speaking ‘good’ or ‘bad’ Chinese or English or whatever.

- A second way to enhance naturalness is to record as lengthy a conversation as possible so that when the conversation is transcribed, the transcriber can select the most natural segment of speech from a much lengthier speech sample. The initial part of a conversation can be discarded, since people are sometimes nervous and hesitant upon first being recorded but after a while become less self-conscious and start speaking more naturally.

- When it comes time to actually make recordings, whoever is making the recordings needs to follow a few basic principles to ensure the most natural recordings possible.
- As far as possible, those making recordings should try to record individuals in the actual environments in which they typically speak. Because the goal is to record people in natural speaking environments, it is best for the research assistant not to be present during recordings. He or she can simply set up the recording equipment, turn on the recorder, and leave. Alternatively, the people being recorded can be loaned the recording equipment and taught to set it up and turn on the recorder themselves.

- While collecting natural speech is a key issue when recording multi-party dialogues, it is less of an issue with other types of speech. For instance, those participating in radio and television broadcasts will undoubtedly be conscious of the way they are speaking, and therefore may heavily monitor what they say. However, heavily monitored speech is ‘natural’ speech in this context, so this is precisely the kind of speech one wants to gather.

- In recording any kind of speech, the type of recorder to use is an important issue to consider. With the rise of digital technology, there are now digital tape recorders that can make high-quality digital recordings.
- When working with spoken recordings, it is often desirable to work with digital recordings on a computer to prepare the recording for inclusion on a CD for storage or distribution.

- Advantages and disadvantages of tape-recorded data (rather than observational data)
- As Erikson (1986)points out, you cannot interact with a tape unless you are part of the interaction being taped, you cannot ,for example, stop people and ask them why they said what they just did, or get them to repeat something that is not clear. In addition, contextual information that might be important may be unavailable. Audiotape does not reveal whether people were looking at one another or how their bodies were aligned; even videotape shows only what the camera was aimed at. It is, of course, possible to alleviate these problems by being there when the tape is being made and taking good notes, but even so there are bound to be questions that arise later.

- On the other hand, Erikson says, tape-recording can help make an analysis more complete and more fine-grained than purely observational approach would. A tape you can hear or see over and over makes it possible to notice and take seriously things that happen only once or rarely. In real-time observation we tend to notice things that are repeated. And a tape can help keep the analyst from jumping to conclusions that may be reinforced by selective memory. Baugh(1993)points out another advantage of tape-recording: the researcher does not have to be the one to collect the data, if he or she is not the ideal fieldworker for the situation.

- **Other data sources**
- There are some kinds of spoken data which you don’t have to record (or ask informants to record) because they already exist and in the public domain. An obvious example is talk that is **broadcast on radio or television.**

- The new medium of the **internet** is another potential source of data, in two ways. First, communication on the net--- on discussion lists, bulletin boards, chat rooms---is an object of considerable interest to discourse analysts, who point out that although the medium is writing, it has many ‘speech-like’ features. Second, there are sites where a user can download linguistic data that has been collected by other researchers. A collection of linguistic data is known as a corpus, and the availability of more and more corpora on-line is an important recent development.

- One obvious drawback with getting spoken language data from an on-line corpus is that typically you only get the written transcript. (Sometimes you can get sound, but it is more likely to be isolated syllables or words than whole chunks of conversation.) However, the transcript may include additional information by way of ‘tagging’, where the original analyst has coded the data, using symbols to indicate things like pitch, stress and volume. If you know the relevant symbol you can search for the phenomenon it codes.

- Whether this kind of data is useful depends on what you want to investigate. Potentially, there is an enormous advantage in having access to corpora far larger than you could collect and transcribe yourself: it is especially advantageous if you want a lot of examples of a very well-defined feature, or if you want to check how common something is statistically.

- **How much data?**
- How much data is enough? It depends on your goals, your resources, and the kinds of claims you are hoping to be able to make.
- Some researchers have worked on very small and specific data samples: Deborah Tannen is famous for an extremely detailed study of the talk which took place at a single Thanksgiving dinner (Tannen ,1984). Jennifer Coates's study *Women Talk: Conversation between Women Friends*(1996) is based on a larger sample of data from several different groups of women friends, reflecting the fact that Coates was interested in similarities and differences among women of different ages and backgrounds.

- A factor that affects how much data you may have to record is how frequently the variable you are interested in occurs in talk. For instance, there is a well known article called " On questions following questions in service encounters" (Merritt, 1976). Questions following questions in service encounters (for example 'Can I have a pint of Guinness?' 'Are you 18?') are not exactly rare, but nor do they occur in every service encounter: a researcher will need to record quite a large number of encounters to get a reasonable number in which questions follow questions.

- By contrast, if you are researching something like hesitations, or minimal responses like ‘mhm’, ‘yeah’, you can be confident that they will occur very frequently in any exchange you record. It will not take long to collect a respectable number of examples.

- While **in principle** it might seem desirable to have as many examples as possible, **in practice** you have to draw the line somewhere, bearing in mind that the collection of data is only the beginning of a lengthy process which includes **transcription** as well as the actual analysis.

Transcribing the Data

- Transcription involves capturing who said what, in what manner (e.g. prosody, pause, voice quality), to whom, under what circumstances (e.g. setting, activity, participant characteristics and relationships to one another). It includes preservation of various temporal aspects (e.g. pause duration, sequence of events, and simultaneity or overlap of speaker turns or speech and gestures), and some metacomments.

- Transcribing speech is in essence a highly artificial process, since an exclusively oral form of language is represented in written form. Consequently, before any transcription is undertaken, it is important to decide just how much that exists in a spoken text one wishes to include in a transcription of it. In the research done in the past, researchers have varied considerably in how much detail they have included in their transcriptions of speech.

- Whether one does a minimal or detailed transcription of speech, it is important to realize that it is not possible to record all of the subtleties of speech in a written transcription. As Cook(1995) notes: a spoken text is made meaningful by more than the words one finds in a transcription: how a conversation is interpreted depends crucially upon such contextual features as paralanguage (e.g. gestures and facial expressions), the knowledge the participants have about the cultural context in which the conversation takes place, their attitudes towards one another, and so forth.

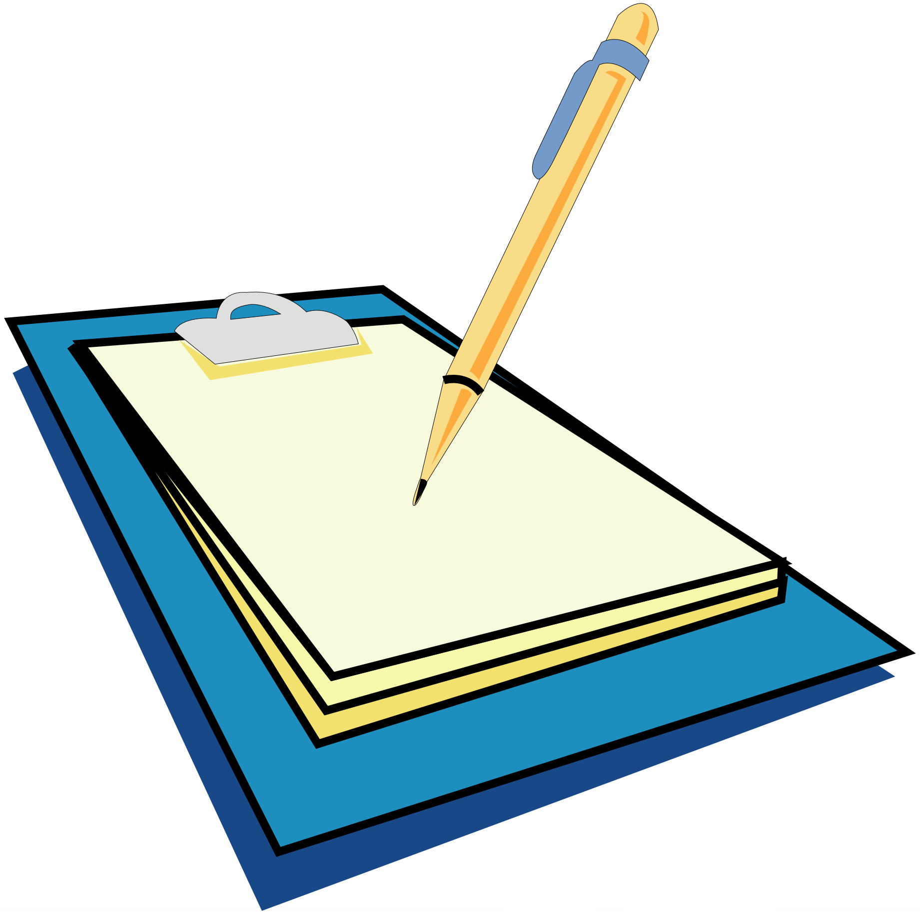
- All of this extra-linguistic information is very difficult to encode in a written transcription without the researcher developing an elaborate system of annotation to mark this information and the transcriber spending hours both interpreting what is going on in a conversation and inserting the relevant mark-up. It is therefore advisable when transcribing speech to find a middle ground: To provide an accurate transcription of what people actually said in a conversation, and then, if resources permit, to add extra information (e.g. annotation marking features of intonation) A few simple steps that will make the transcription easier to read and refer to:

1. Use actual, spelled out names for the speakers, not initials.(These should generally not be their real names, but rather pseudonyms)

2. Number the lines of the transcript so you can refer the reader to the precise parts you discuss in your paper.

3. In excerpts from transcripts that you use as examples, highlight the relevant lines, words, or phrases so readers can pinpoint exactly what to focus on.

Design Your Own Project



- **Choosing your research topic**

  Most ‘good ideas’ for research do not just spring from the researcher’s imagination, they are suggested by previous research. Knowledge on any subject is cumulative: people look at what is already known and notice gaps, or think of objections or supplementary questions. There are several ways in which previous research can be used to help design a new research project.

- **Replicate**

  The simplest kind of research project based on existing research involves replicating a previous study. You ask the same question(s) and use the same methods as a researcher whose work you have read, but you collect your own data: your question is whether your findings will resemble the original findings. This might sound dull, but in fact it is quite important to know whether patterns reported in research turn up consistently. Only when similar results have been obtained across a range of different studies can we make reliable generalizations about whatever is at issue. Another reason for replicating old studies is to find out if things have changed over time.

A variant on straightforward replication is to undertake a study asking the same questions and using the same methods as a previous study, but deliberately altering one or more of the nonlinguistic variables(e.g. the social characteristics of the participants or the setting for their interaction).

- **Compare and contrast**

  Another common type of research project compares two kinds of talk. E.g. a comparison between spontaneous casual conversation and soap opera dialogue; a comparison between face-to-face talk and telephone conversation etc.

- **Take issue with a previous claim**

  Sometimes you may feel intuitively that a claim is misguided and then try to demonstrate empirically why it is misguided.

- **Describe something new**

  You can also construct a worthwhile research project by focusing on a kind of data that has not been analyzed before, and simply describing its discourse characteristics. For instance, Kay Richardson(1997) has analyzed the talk that is heard on 'shopping channels' that work like mail order catalogues, offering goods that viewers can buy over the phone. Shopping channel presenters face a linguistic challenge: since viewers need to be able to inspect the object on offer for long enough to decide if they want to buy it, presenters have to talk about a single object , like a ring or a watch, for much longer than would be typical in other genres of talk. They have to describe the object, with particular attention to features the camera might distort. They also have to make it sound attractive, and constantly repeat information like the price and the phone number to call if you wish to order. The resulting discourse is very distinctive.

How is previous research relevant to a project that involves describing something new? In two ways. First, it is desirable to investigate whether anyone else actually has published a description of your chosen genre. Second, there may be a case for approaching a novel phenomenon on the model of something that has already been studied.

What makes a workable research topic?

- A well-focused idea can be phrased as a question or a set of closely related questions. (Your questions may become narrower in the course of your research, but you must have clear questions with which to start.)
- You know how you will go about answering the questions, that is, what your research methodology or methodologies will be.
- You are familiar with the research site.
- You have the time it will realistically take to carry out the research, analyze the results, and write your report.
- The project requires ways of collecting and analyzing data with which you are comfortable and at which you are competent.

- **Putting it in writing**

  Most research papers on spoken discourse will have some variant of the following structure:

1. **Introductory** material setting out the researcher's question and why it is of interest.
2. A **literature review** surveying relevant previous work on the subject.
3. A **methods** section explaining in some detail what the researcher's data consist of, how they were collected, what approach was taken to analyzing them and why that approach was chosen as most suitable for the purpose.
4. A section setting out the **results** of the project

5. A **discussion** section in which the researcher presents possible explanations for the results and discusses their meaning or significance.
6. A **conclusion** in which the researcher summarizes what has been learnt from the project, what its problems and limitations are, and what questions arise from it for future research.
7. A **bibliography** listing all the resources the researcher has consulted
8. **An appendix** setting out transcription conventions used by the researcher.
9. **An appendix** containing transcribed data.

Thank you !