PhD Course in Advanced Sociolinguistics

-Data Collection and Research Question

- Questionnaires and Interviews

Prepared by Musaab N, Muhammad

Supervised by Asst. Prof. Dr. Wafa'a M, Ali

Data Collection

Data collection is a systematic process of gathering, measuring, and recording information or facts from various sources to gain insights, analyze patterns, and make informed decisions. It forms the basis of research, business operations, and virtually any field where information is used to understand phenomena, evaluation purposes or drive decision-making. Data collection in sociolinguistics is a fundamental step in studying how language functions within social contexts.

Sociolinguists aim to understand the interplay between language and society, so collecting accurate, rich, and relevant data is essential for meaningful analysis. This process must be approached critically to ensure validity, reliability, and ethical soundness. Sociolinguistic data refers to linguistic behavior observed in real-life social settings. It includes spoken or written language, often accompanied by metadata such as the speaker's age, gender, ethnicity, education, occupation, and social class. The primary aim is to investigate how these social variables influence language use, variation, and change.

Data collection is critical for ensuring data integrity, as flawed information—whether inaccurate, corrupted, or incomplete can contaminate every initiative undertaken with that data. The quality of collected data directly impacts the validity and reliability of any conclusions drawn from it.

Data collection involves gathering numerous data points from different sources to transform them into actionable information. The goal is to retrieve information that can be utilized to recognize patterns, improve processes, test hypotheses, or answer research questions. Data can be collected from multiple channels including surveys, focus groups, interviews, questionnaires, observations, existing databases, web pages, and more. It can be automated (like web scraping) or manual depending on the nature of the data and research objectives. Data can be either expressed in numbers and graphs, analyzed through statistical methods or expressed in words, analyzed through interpretations and categorizations.

The Data Collection involves the following processes:

1. Preparation Phase: Define the aim of your research or analysis

=Identify the type of data needed (quantitative or qualitative)

=Determine appropriate collection methods and procedures

=Allocate necessary resources (personnel, technology, materials)

=Train data collection teams on instruments and protocols

=Address ethical considerations and obtain necessary permissions

2. Implementation Phase: Follow established protocols consistently

=Maintain supervision of collection activities

=Document the process thoroughly, including dates and locations

=Implement data security measures to protect collected information

3. Post-Collection Phase: Organize data into appropriate formats (tables, charts, etc.)

=Store and manage collected information securely

=Ensure accuracy and quality of collected data

=Process data for analysis and interpretation

Research Questions in Sociolinguistics

In sociolinguistics, research questions are central in guiding investigations into the relationships between language and society. These questions are carefully formulated to address specific aspects of linguistic behavior, language variation, and the social factors that influence them. A strong research question in sociolinguistics is clear, focused, researchable, and grounded in theoretical or empirical issues within the field.

Understanding the history of sociolinguistics will help students to ask better sociolinguistic research questions. Scholars who work in the field of sociolinguistics come from different academic backgrounds, and they pose different kinds of research questions. This seminar illustrates how some scholars ask research questions more focused on language, while others ask research questions more focused on society. Both areas of study serve to further the goals of sociolinguistics, but the researcher must choose one as primary in order to create a focused and coherent research project.

Sociolinguistics has been a diverse academic field: scholars from linguistics, anthropology, and sociology came together because of their collective interest in the study of language in its social context **(Bright, 1966).**

- • The linguists used information about society to better explain how language works,
- • while the sociologists and anthropologists used language variation to better explain how society works.

Research questions are important because they guide the researcher's time, and time is a valuable, vanishing resource. Empirical data from both language and society are used in many studies, but students need to decide which way they are going to tackle in their research projects.

The choice of research question determines the kind and amount of data one may need, and it determines the need for qualitative methods only or for quantitative methods as well.

What Sociolinguists favor? Sociolinguists greatly favor language resulting from human interaction (versus data constructed by linguists themselves). Such language is open to a multitude of analysis methods to achieve many different research goals.

The technology used within sociolinguistic studies has become much more sophisticated over the last four decades in some ways, the results of these changes should be very obvious. In the 1960s,

What makes a well-designed research question?

1. It should be based on previous research.

2. The research question should extend the knowledge of the field in some way.

3. It should be practical. Make a research question that is doable in a given time frame.

4. It should be simple. Everyone should easily understand where the project is headed.

For methodologies focused on social questions, there have been similar improvements in collecting data (e.g., web- based surveys), but the most notable changes to those research questions involve refined definitions of the objects of study. For example, whereas early work focused on how women and men speak differently, later work focused on how people use their sociolinguistic resources to construct gender.

Types of Sociolinguistic Research Questions

Descriptive Questions – Seek to document and describe variation.

Example: "What are the phonological features of African American Vernacular English (AAVE) in Detroit?"

Purpose: To create an inventory or map of language use within a community.

Comparative Questions – Compare language use across groups or contexts.

Example: "How does code-switching differ between bilingual Spanish-English speakers in formal vs informal settings?"

Purpose: To identify patterns and differences influenced by context.

Causal/Explanatory Questions – Seek to explain why certain linguistic behaviors occur.

Example: "Why do younger speakers in urban areas adopt more innovative slang terms than older generations?"

Purpose: To investigate social or psychological motivations for change.

Critical Questions – Investigate power, ideology, and inequality in language use.

Example: "How does institutional language in legal settings marginalize non-native English speakers?"

Purpose: To expose hidden power structures and advocate for social justice.

Application of Research Questions

To a growing extent, changes in research questions have developed in terms of where the variationist methodology is applied. Variationist research questions have been applied to previously under-researched languages, such as sign languages. With the linguistic components of sign language, such as phonology, language variation patterns have been found to operate much as they do with spoken languages, demonstrating variability and the influence of social factors. Some of those research questions, focus more on linguistic questions while others focus more on social questions and personal identity.

Social Research Questions

Most research studies examine language variation to learn more about social factors relating to societies and individuals.

Development of Research Questions

1- Earlier research questions focused on topics such as **race** and **sex**, while

2- later ones examine the mutual influences from areas such as **ethnicity** and **gender**.

3-Now, sociolinguistic studies regularly examine **style, identity**, and **social meaning** through language analysis. This subsection details some of the research questions that have been asked in these pursuits and examines how they have changed.

Social analysis in sociolinguistics has seen dramatic changes since the 1960s.

- In the early days, the key method was
to correlate demographic categories and linguistic variables. In most of those studies, the goal was to figure out how the social factors influenced the language variation patterns under study. This broad correlation technique is still a method used to assess dialect regions and language change in larger communities, but it is mainly employed by variationists to answer linguistic questions.

o The range now reaches from broader levels of society
1-to social networks with different levels of density and multiplexity.
2- to communities of practice
3- down to the individuals who contain a model of the entire social macrocosm in their heads and who (re)create sociolinguistic styles in the ebb and flow ( )المد
والجزر of social meaning and personal identity.
An important change for social research questions is the object of study itself.

The difference between early and modern scholars

- **Early scholars** investigated monolithic (harmonized) categories like race (e.g., Black, White) and sex (e.g., women, men),
- **Modern research** questions explicitly discuss the natural complexity of social constructions like gender and ethnicity. In addition to sexual orientation factors which went unexamined in the early days for several cultural reasons but these factors are now an essential part of sociolinguistic research.

Topics included in Sociolinguistic studies through time

Sociolinguistic Research of Ethnicity

Numerous other social areas have undergone similar transformations over the last 40 years. Early studies of ethnicity were sociolinguistic descriptions of the language variation patterns of various ethnic groups. For example,

1- Wolfram (1969) examined the dialect of African Americans in Detroit. and, then, the dialect of Puerto Ricans in New York (1974).

2-Labov *et al*. (1968) examined the language variation patterns of African Americans and Puerto Ricans in New York City.

3-Fasold (1972) did the same for African Americans in Washington, DC.

Features of Linguistic Variables in Ethnic Groups

1-They were front and center in these studies.

2-The results reflected differences and similarities between and within ethnic groups.

3-The research questions in such studies just as the following examples:

How do African Americans speak differently from other ethnic groups?

At what rates do these variable rules operate for different social classes?

Awareness and respect for ethnic diversity was a fundamental part of all these studies, but ethnicity itself as a social construct was not the focus of study. The study of ethnicity and many other social factors was enhanced by increasing attention to the interactions of social factors.

- **Sociolinguistic Research Studies of Gender**

Kirk Hazen maintained that: with ethnicity, early variationist studies focused on how females and males used linguistic variables differently.

The contrast between the earlier days and more modern studies is that:

The terms “sex” and “gender” have been recognized as separate for decades.

- **Sociolinguistic Research Studies of Identity**

Research questions of sociolinguistics examine style in the sociolinguistic construction of identity, and the basic assumption is that social categories like ethnicity, gender, and sexual orientation are co-constructed between speakers and audience.

Later studies of early variationist tackled:

1. Dialectology
2. Sociology projects and took up demographic social descriptors.
3. larger sociological categories and the speech community to model language change.
4. What language variation patterns mean for individuals, the answers from percentages and linear regressions of demographic categories and phonological conditioning environments did not speak as loudly as many scholars would have liked.
5. The fields of anthropology and social psychology.

The essentials

Starting in the middle of the **twentieth century**, many branches of the humanities began to analyze and question arguments which assumed that social groups had distinct foundational qualities.

- Was it part of a woman's essential nature to wash more dishes or talk more (or less) often than a man?

-Were differences between ethnic groups a result of their essential natures?

- Do social groups even have an essential nature?

Scholars began to question and refute these ***essentialist*** claims.

QUESTIONNAIRES

According to Rasinger (2008: 57), questionnaires today are “the most popular tool in quantitative methodology” in that they help to “accumulate vast amounts of high-quality data”.

Milroy & Gordon (2003: 52) add that questionnaires are a very efficient means of collecting data: “they allow researchers to gather data from a large number of speakers in a relatively brief amount of time.” In sociolinguistic research questionnaires are usually used in order to elicit data **about language**, but **not** data **on linguistic performance** (Codó 2008: 171). **Due to this limitation**, questionnaires are often **combined** with other types of data collection, for instance, interviews, participant observation, ethnographic notes, etc.

According to Codó (2008: 171), knowledge obtained from **surveys** can provide you with a **general picture** of language use in a given community; it can also help you to identify target subjects for participant observation or ethnographic research. Thus, **depending on your research questions**, your research design might **rely** just on data obtained from **questionnaire survey**, or **questionnaires** can be just one of several data sources you need to use.

Most **questionnaire** items **consist of** (i) **instructions**, (ii) **a question** and (iii) **possible answers**. **Questions** should be **separated very clearly from the instructions** *using different typefaces or fonts*. **Response options should cover a range of responses for a variety of respondents**.

Types of Questionnaires

Closed Questions

Questions that provide a **closed range of possible answers** selected by the researcher, and respondents are asked to choose one or more of these answers or organize them in a multitude of other ways. There are **different types of closed questions**; the most suitable for sociolinguists are introduced below:

 Checklists: a selection of possible answers, and respondents can select as many as they wish that apply to a particular statement or question, for example:

Q-Which of these do you think speak a widely recognizable local dialect?

A-London Liverpool Leeds Manchester Birmingham

- **Rankings:** an extension of the checklist. They ask respondents to place the options in terms of preference, frequency, or some other variable. For example, respondents could be asked to order the options from the example above in terms of most widely recognizable to least recognizable.
- **Rating scales:** allow respondents to express their degree of agreement with a statement, such as in the scale below:

Q- Some people believe that one should never use more than one language in one sentence. To what extent do you agree with this statement?

Not at all Very

1 2 3 4 5 6

 Semantic differential questions: Informants are asked to select a position on the scale that best describes their feelings.

Q- Listen to the following voice and put a cross on the line where you would put this person on this scale.

A- Reliable Unreliable

 True–false questions when the respondents, like children, are not be able to give complex answers, this is a very appropriate question type. The key sentence has to be **short** and contain **only one main idea**, which must **not be subject to debate**.

Q- Which of the following statements do you agree with? Please circle the correct answer.

A-I think women talk more than men. True False Undecided

B-I have more male than female friends. True False

 Multiple-choice questions: a question or statement and a selection of possible answers.

Q- Based on the recording you've just heard, this person gives the impression of being (check all that apply):

Reliable

Educated

Friendly

Other (please specify): ________________

Open-ended Questions

In this type of question, respondents do not select from pre-formulated answers but provide their personal answers. This may be a more precise and personal response than those given to questions of the closed type.

There are different types of open-ended questions:

- **Specific open-ended questions** ask about a precise piece of information.

- **Questions of clarification** often follow specific closed questions when a particular answer was chosen; for example, *If you believe English should be the only language medium in school, could you explain why you believe this to be a good policy?*

- **Sentence completion items** ask respondents to complete a sentence that should point to a well-defined issue; for example, *one thing I like about bilingual education is …*

- **Short-answer questions** ask for one concept or one idea; for example, *please make two suggestions for how bilingual education could be improved at your school?*

Structuring the Questionnaire

The standard structure of a questionnaire consists of: Introduction, Mid-section, conclusion.

The introduction

In this section, the questionnaire introduces the research to the respondents. It should minimally include:

1. Title of the questionnaire.

2. a *brief* explanation of the purpose of the research and who is responsible for conducting the study.

3. a polite request to fill in the questionnaire fully and honestly.

4. a short outline of what the questionnaire will cover and how long it will take to complete.

5. a promise of anonymity and confidentiality.

6. the researcher’s name, institution, and contact details.

7. an expression of thanks.

The Middle

In this section, the questionnaire comprises the questions which should be a clear logical structure. Similar topics should be grouped together. There should also be question instructions that explain how respondents should answer. They should explain and exemplify what rating scales there are and how they work. Questions should be factual, interesting and clearly relevant to the research topic to motivate the respondents to complete the questionnaire. Besides, a questionnaire should not take more than half an hour to complete.

The Concluding

The concluding section should include:

1. Expression of thanks for taking the time to complete the questionnaire.
2. Contact details, in case respondents have questions about the research later.
3. Renewed promise of anonymity and confidentiality.
4. Ask the respondents to check all questions have been answered.
5. Describe how questionnaires should be returned (if this is not clear).
6. Explain how survey results can be obtained (e.g., via a web site) or promise feedback (in the form of a poster, a meeting, or an article).
7. Invite respondents for a follow-up interview.

Testing the Questionnaire

A draft version of the questionnaire should be tested thoroughly before it is used on the target participants for the following purposes:

1. to ensure sufficient information is provided to the participants and instructions are clear.
2. to ensure all questions are understood and answerable in the intended sense.
3. to ensure the questionnaire is a reasonable length, and participants are motivated to complete all questions.
4. to ensure the elicited data will be valid, complete, reliable, and can be analyzed efficiently.

Phrasing, Lay-outing, Numbering

We should phrase our questions carefully by avoiding the following:

- Academic jargon or too much of linguistic terminology and we should adjust the questions to the level of understanding of respondents.
- Not mentioning number of responses will be accepted per question. This is very important for multiple-choice answers: in some cases, we might need just one answer, in other cases, several answers can apply.
- Asking two questions in one, this will confuse your respondent.
- Long and complicated questions.
- Leading questions like "*Do you think knowing more than one language is an advantage?*".
- Ambiguity: use less abstract nouns and adjectives.

Sociolinguistic Interview

Wolfram (2010: 302) calls the sociolinguistic interview "the methodological heart of the sociolinguistic movement". It is used by sociolinguists to investigate language use in relation to social factors such as age, gender, ethnicity, social class, and culture.

Milroy & Gordon (2003: 60) say that "the general idea of the interview is to use interlocking modules to simulate the seamless topic-shift structure of normal conversation."

Labov (1984: 37), “the most successful interviews follow the path which is both natural to the speaker and comparable to other paths”. While unsuccessful interviews are those, during which the speakers only answer questions. The interview should simulate a natural conversation. It can be a difficult task to accomplish, especially if you are interviewing a complete stranger. The interviewer should pay attention to the topics where the respondent feels the most comfortable and confident. Elaborate on these topics and be flexible. This will help to elicit good data.

Parts of Sociolinguistics Interview

According to **Meyerhoff** et al. (2012) the “classical” sociolinguistic interview consists of at least **four** parts:

A. Reading a **list of minimal pairs** (word pairs, e.g., *pin* and *bin*).

B. Reading **a list of words in isolation**.

C. Reading **a short narrative**.

D. Talking with the interviewer.

The purpose of the first three is to elicit a continuum of different styles for every speaker (Meyerhoff et al. 2012).

Goals of Sociolinguistic Interview

One of the most important goals of sociolinguistic research is to obtain "natural speech" used in ordinary, everyday interactions with all the variability that entails (Meyerhoff et al. 2012: 123). Sociolinguists are interested in the "real" language people use. So, the vernacular has been an object of investigation since the very first sociolinguistic studies. One of the most often used techniques to elicit the vernacular is the sociolinguistic interview. However, Labov (1984) lists ten of them:

1. To record at least one to two hours of speech of each speaker.
2. To obtain a full range of demographic data (social class, age, gender, school, occupation, income, group membership, and other).
3. To obtain comparable responses to questions that define different experiences and attitudes (such as the famous question of the danger of death).
4. To elicit narratives of personal experience (that will reveal community’s attitudes and norms toward particular phenomena).
5. To stimulate group interaction during the interview (so that the attention is deviated from the interviewer).

6. To allow the interviewee lead in defining the topic of conversation (this might result in longer narratives by the interviewee and less-monitored speech).

7. To trace the patterns of communication among members of the speech community.

8. To obtain overt attitudes toward language, linguistic features and linguistic stereotypes.

9. To obtain linguistic information through formal elicitation: reading text and word pairs.

10. To carry out field experiments on subjective evaluations toward perception of linguistic forms.

According to Labov, to achieve these goals, a sociolinguist has to use two technical devices:

1. **Conversation Module**

A group of questions focusing on a particular topic like *childhood experiences*, *university experiences* and other (Labov 1984: 33). They **help to construct the structure of the interview**. Labov relied on a set of the module structure, called Q-GEN-II in a research in Philadelphia speech community: Module1: Demography; Module9: Family; Module5: Marriage; ……etc.

2. **Conversational Network**

All of the modules must be combined into a conversational network. Modules begin and end with *transitional questions* that connect to another module in the conversational network by the interviewer or might occur naturally. For example, module 2 “Games”, might lead to the questions associated with module 3 “Fights” or module 11 “Peers”; the latter module might lead to module 4 “Dating”, etc. (Labov 1984: 34)

Observer’s Paradox

It is an “inescapable fact that speakers are more aware of what they are saying and how they are talking as soon as you begin recording them” (Meyerhoff et al. 2012: 123) that a phenomenon every sociolinguist conducting an interview or observation has to consider.

Meyerhoff et al. (2012: 124–125) point out several methods that help to mitigate the effects of the observer’s paradox:

1. Modification of the number of people in an interview. Most people feel uneasy when they are interviewed in a one-to-one setting. Thus, inviting several people to participate in a conversation might result in better data and a relaxed interview.

2. Modification of the number of interviewers. This may not seem usual, However, interviewers might frame themselves as friends and their natural conversation might develop into a less-formal interview with the respondent.

3. Removal of the interviewer. This strategy minimizes the effect of outsider presence. However, then there is no control over the recording.

A change in the content of the interview sometimes helps to achieve more casual speech. It has been noticed by many sociolinguists that "when people are emotionally involved (excited, angry, fearful, etc.) in a discussion, they are more concerned with what they say than with how they say it" (Milroy & Gordon 2003: 65). Like "*danger of death*" a question used by Labov. However, this kind of question may not be appropriate to ask in certain situations. If we are conducting an interview with a child and ask whether he ever feared for his life, we may not receive an expected answer. On the other hand, not all people feel comfortable remembering painful experiences of their lives.

Type of Interviews

Semi-Structured

The “classical” sociolinguistic interview (used by Labov and Trudgill) is a semi-structured interview in which, researchers prepare questions in advance and during the interview they can move from one question to another. It also “allows the researcher to offer topics that participants can talk about and lead them to have more dynamic discussions between themselves” (Meyerhoff et al., 2012). Thus, the researcher does not take a full control over the conversation and this helps to create a more relaxed atmosphere during the recording.

The drawback of such a technique is that some people might digress to some topics, while others to other topics and the data may not be comparable between respondents. On the other hand, we may end up having a large amount of natural speech data.

Structured Interviews

In structured interviews (e.g., *questionnaire-based*) participants do not self-select questions for further discussion. Such interview has certain drawbacks, since usually it is hard to achieve a relaxed atmosphere during the interview. On the other hand, a researcher has the control over the interview and elicits all of the required information. If we interview many people using structured interview, this data becomes comparable between respondents.

Questioning Techniques

Sometimes it might seem impossible to elicit answers from someone who is not willing to collaborate. So, there are some tips provided by Tagliamonte 2006:

 Tailor your style to suit your interviewee, modify your language if needed, but do not overdo it.

 Avoid yes or no questions. In other words, do not start questions with “Do you”, but rather, What?, Why? and How?

- Use indirect means, e.g., “Is it true that...”, “I’ve heard that...”.
- Be careful when asking about person’s age, education, political views, religion.
- Word your questions in terms of other people (in general), rather than asking something specific about the interviewee (“Some people say that it was much better to live during Soviet times. What do you think?”).
- Do not ask too personal questions (how much money do you make, what are your religious beliefs, etc.).
- Organize your questions so that one question naturally leads to the next, prepare your questions in advance.

- Let the informant talk, do not interrupt the interviewee.

- Use your natural variety (do not sound too formal or too standard).

- Relax and speak informally (it is not easy to do!).

- Ask short questions, do not formulate long, complex sentences.

PhD Course in Advanced Sociolinguistics

Language and Gender

Prepared by Musaab N, Muhammad

Supervised by Asst. Prof. Dr. Wafa'a M, Ali

1-Introduction:

Over the past thirty years, and as a result of the women's movement, gender issues have become connected with the issue of language. Gender studies and language studies are interdisciplinary academic field. The study of language began from thousands of years ago, while the study of gender is quite short.

Gender studies have developed differently, achieving the greatest influence in North America; the era of feminism that began in the late 1960s and affected academic and public life as well as high and popular culture has been instrumental in shaping the historical and scholarly context of its generation. Feminist work has entered and had an impact upon almost every academic discipline. (Flotow, 2004:1) this movement crystalized in production many work concern language and gender as following:

- George Keith and John Shuttleworth (2008: 222) in ‘Living Language’ suggest that: women talk more than men, talk too much, are more polite, are indecisive/hesitant, complain and nag, ask more questions, support each other, are more co-operative while men swear more, don't talk about emotions, talk about sport more, talk about women and machines in the same way, insult each other frequently, are competitive in conversation, dominate conversation, speak with more authority, give more commands, interrupt more.

Deborah Tannen’s views:

- Physical orientation:male avoid eye contact while female use eye contact.
- Status and connection: male talks for status and female talks for solidarity.
- Directness and indirectness: male generally moves from decision to discussion in contrast female moves from discussion to decision. Male is talkative in public and quite in private while female is quite in public and talkative in private.
- Jennifer Coates (1993) claims that: Men will often reject a topic of conversation introduced by women while women will accept the topics introduced by men, they discuss ‘male’ topics e.g. business, sport, politics, economics. Women are more likely to initiate conversation than men, but less likely to make the conversation succeed.

- Robin Lakoff in her book ‘Language and Woman’s Place (1975) and in a related article published some claims that women speak less frequently: show they are listening by using minimal responses mm, yeah; Speak more quietly than men and tend to use the higher pitch range of their voices; Use hyper-correct grammar and pronunciation: Standard English, use a greater range of intonation and ‘speak in italics’: so, very, quite, use question intonation in declarative statements, women make declarative statements into questions by raising the pitch of their voice at the end of a statement, expressing uncertainty;

Overuse qualifiers: (for example, “I think that...”); Hedge: using phrases like “sort of”, “kind of”, “it seems like”; Use super-polite forms: “Would you mind...”,“I'd appreciate it if...”, “...if you don't mind”; Apologise more: (for instance, “I'm sorry, but I think that...”); use tag questions: “You're going to dinner, aren't you?” Have a special lexicon: e.g. women use more words for colours, men for sports.

2. What is Gender?

Gender is the state of being male/masculine or female/feminine in relation to the social and cultural roles that are considered appropriate for men and women (Online Collins Dictionary). It’s not the same as sex. Sex refers to the biological characteristics of women, while gender is determined by the conception of tasks, functions, and roles attributed to women and men in society and in public and private life. That is to say, gender is socially constructed. **In grammar,** the gender of a noun, pronoun, or adjective is whether it is masculine, feminine, or neuter. A word's gender can affect its form and behaviour. In English, only personal pronouns such as 'she', reflexive pronouns such as 'itself', and possessive determiners such as ‘his’ have gender.

- **Semantically,** gender was associated with the differentiation of sex. Therefore, the formal grouping into genders was smoothly and naturally superseded by a semantic division into inanimate and animate nouns with a further subdivision of the letter into males and females.
- In **sociolinguistics** and other social sciences, gender refers to sexual identity in relation to culture and society. How words are used can both reflect and reinforce social attitudes toward gender (cf. Eckert, 1998).
- Talbot (1998: 7) points out that people acquire characteristics which are perceived as masculine and feminine. Thus, according to Talbot (ibid), gender is learned. People normally talk about a masculine woman or a feminine man, i.e. a person can be more or less feminine and more or less masculine (Graddol & Swann, 1989: 8).

Furthermore, gender is not binary; we can talk about one man being more masculine (or feminine) than another. There are, grammatically speaking, comparative and superlative forms of masculine, i.e. more ‘masculine’ and most ‘masculine’, but it is not the same with ‘male’ (there is no ‘maler’ or ‘malest’). Talbot (1998: 7) states that: people are gendered, and actively involved in the process of their own gendering. In this way also, it is clear that sex-exclusive and sex-preferential differentiations are part of doing gender and of behaving as proper men and women in particular cultures (ibid); they are cultural in essence (and not biological) since they differ from one culture to another.

The **gender approach**, as Wijk & Francis (1999: 1) argue, is distinct in that it focuses on women and men, and not on women in isolation. It highlights the following:-

1. The differences between women’s and men’s interest even within the same household and how these differences interact and are expressed.
2. The conventions and hierarchies which determine women’s and men’s position in the family, community, and society at large, whereby women are usually dominated by men.
3. The differences among women and among men, based on age, wealth, ethnic background and other factors.
4. The way gender roles and relations change, often quite rapidly, as a result of social, economic, and technological trends.

3 Language and Gender

Language is a system that constitutes meaning. The question of language and its political implications has exercised writers, philosophers and social theorists throughout the intellectual history of western civilization (Cameron, 1985: 1).

Early language and gender research focused on documenting empirical differences between women’s and men’s speech, especially cross-sex interaction; describing women’s speech in particular and identifying the role of language in creating and maintaining social inequality between women and men.

The goal of uncovering the role of language in maintaining gender inequality is evident in the field's foundational text, Robin Lakoff's *Language and Woman's Place*, one of the first to call attention to gender differences in ways of speaking. Lakoff describes "an attempt to provide diagnostic evidence from language use for one type of inequity that has been claimed to exist in our society: that between the roles of men and women".

Consequently, the language and gender field consists of two main aspects. The first settled as part of quantitative sociolinguistics: sociolinguists analysing the co-variation of language and variables such as social class began to notice that their data also revealed gender differences, i.e., the focus is on phonological, morphological, or lexical features of language.

For example, Peter Trudgill (1974, 1983), examining the pronunciation of a wide range of speakers living in Norwich, UK, realized that women and men of the same social class are patterned differently. Women on average used forms closer to **Standard English**, while male speakers used a higher proportion of **vernacular forms**. Trudgill's analysis demonstrates that the use **of non-standard forms** of language seems to be associated not only with working-class speakers but also with male speakers and thus with masculinity. This strand continues to flourish, with more recent research taking a community of practice approach.

The second aspect of language and gender research focuses on language as a ‘concrete living totality’ (Bakhtin, 1981), in other words, on sociolinguistics**.** The move in linguistics from the micro-analysis of phonemes and syntactic structure to a more macro-analytic approach, looking at language more holistically, was undoubtedly a paradigm shift with significant consequences. **The freedom to think about talking in general and analyse whole conversations has led to a new understanding of the relationship between language and social life.** A huge emphasis was placed on using authentic language data and analysing them in their social context.

4- Approaches to Gender

1.4 The Dominance (status/power) Approach:

Male speech is dominant over female speech since men are more likely to use power. Power is of two types: force and soft power (to control others’ actions and thinking indirectly via persuasion. Language is one of the most powerful tools which allows men to express their role in establishing the social norms which cause inequality in the behavior and treatment between men and women. Due to such power, men do most of the topic, i.e. they have control over the topic.

Although such approach doesn’t maintain that women speech is not acceptable, but it reflects the dominance position of men in the society through interaction with women. It is clear that women are less concerned with power in the sense that they are satisfied with their subordinate status. They are more concerned with maintaining social relationships. Whereas men are more concerned with power and they want to be leaders.

Thus, women in their communication tend to focus upon building and maintain relationships, on the one hand, men reflect their desire to maintain status and power in relationships.

2.4 The Difference (Cultural) (Socialization) Approach

This approach stresses the differences between the language used in males and females and therefore to construct female and male gender identity.

Boys and girls are brought up differently. They are injected by different ideas and behaviors by elder people who taught them to do so and not to do so, i.e. they plant differences in their children depending on their sex (whether males or females).

Thus, separate groups based on their sex appear and in the course of this they develop contrasting linguistic habits that underlie subsequent miscommunication between them. On one hand, groups of girls do not tend to be hierarchical; decision are taken in common, with the use of suggestion structures, the inclusive “we” and epistemic modality (e.g. may, might, perhaps, etc.). on the other hand, groups of boys are seen as socializing in gangs which are hierarchical in which competition is present. In general, male interactions are characterized by displays of power; female communication by display of solidarity.

3.4 The deficit Approach`

This approach is linked to the linguist *Robin Lakoff* and her influential work *‘Language and Women’s Place‘*. In this study, Lakoff identified several differences in the way women used language when compared to men. Among these are claims that women use:

- **Hedges:** using phrases like “sort of”, “kind of”, “it seems like”, and so on.
- Use (super)**polite forms**: “Would you mind...”, “I'd appreciate it if...”, “...if you don't mind”.
- Use **tag questions**: “You're going to dinner, aren't you?”
- Speak in **italics**: intonational emphasis equal to underlining words - so, very, quiet.
- Use **empty adjectives**: divine, lovely, adorable, and so on
- Use **hypercorrect** grammar and pronunciation: English prestige grammar and clear enunciation.
- Use **direct quotation**: may, paraphrase, more and often.
- Have a **special lexicon**: women use more words for things like colours, men for sports.

- Use **question intonation** in declarative statements: women make declarative statements into questions by raising the pitch of their voice at the end of a statement, expressing uncertainty. For example, “What school do you attend? Eton College?”
- Use “**wh-” imperatives**: (such as, “Why don't you open the door?”)
- Speak **less** frequently
- Overuse **qualifiers**: (for example, “I Think that...”)
- **Apologise more**: (for instance, “I'm sorry, but I think that...”)
- Use **modal** constructions: (such as can, would, should, ought - “Should we turn up the heat?”)
- Use **indirect commands** and requests: (for example, “My, isn't it cold in here?” - really a request to turn the heat on or close a window)
- Lack a sense of **humour**: women do not tell jokes well and often don't understand the punch line of jokes.

Lakoff suggested that these differences were part of ‘Women’s Language’ and was general seen as **inferior to men.** The ‘Deficit Model’ refers to how this language use contributes to women’s lower status and weaker position in society.

5. Differences in men’s and women’s speech

1- Women talk more/less than men

According to Cameron and Coates (1985), the amount we talk is influenced by who we are with and what we are doing. They also add that if we aggregate a large number of studies, it will be observed that there is little difference between the amount men and women talk. On the one hand, in a recent study, Dr. Brizendine (1994) states that women talk three times as much as men. On the other hand, Drass (1986), in an experiment on gender identity in conversation found that men speak more than women.

2-Differences in Their Attitudes toward Language

Women pay more attention to use standard language than men do, so they are stricter with the rules of the use of language.
Example: Man: Are you comin‘‘?
Woman: Are you coming?
Women tend to use the standard form. This point is emphasized in the difference of pronunciation.
In Detroit, people like to use multiple negations, such as: I do not want none. Research found that men use much more of this kind of substandard form than women. This can be seen from movie “Forrest Gump. Influenced by the southern accent, Forrest often uses double negative to show negative meaning. (Holmes)

3-Differences in Choosing Topics

In social interaction, men and women have different interests in choosing their topics. When men are talking, they are more likely to choose the topics of politics, economy, stocks, sports, current news. While women have more interest in talking family affairs, such as the education of children, clothes, cooking, fashion, etc. Women’s talk is associated with the home and domestic activities, while men’s is associated with the outside world and economic activities. Thus, while there is a popular prejudice that women talk more than men, empirical studies of a number of social situations such as committee meetings and internet discussion groups have shown the opposite to be true.

Women may talk more in informal occasions than men, but they surely play the second role in the formal occasions and tend to speak less than men. Sociolinguists studied women's silence in public situations as well as the linguistic work they do in their partnerships (Spender, 1980). Besides these differences, other sex-linked differences exist, such as women and men may have different paralinguistic system and move and gesture differently.

4- **Differences in style**

<u>Women break the ‘rules’ of turn-taking less than men</u>

Studies in the area of language and gender often make use of two models or paradigms - that of **dominance** and that of **difference**. The first is associated with Dale Spender (1980), Pamela Fishman (1980), Don Zimmerman and Candace West (1975), while the second is associated with Deborah Tannen (1984). Dominance can be attributed to the fact that in mixed-sex conversations, men are more likely to interrupt than women. It uses a fairly old study of a small sample of conversations, recorded by Don Zimmerman and Candace West at the Santa Barbara campus of the University of California in 1975. The subjects of the recording were white, middle class and under 35. Zimmerman and West produce in evidence 31 segments of conversation. They report that in 11 conversations between men and women, men used 46 interruptions, but women only two. The difference theory was also summarized in Tannen's book *You just don't understand* (1990) in an article in which she represents male and female language use in a series of six contrasts:

A-Status vs. support

This claims that men grow up in a world in which conversation is competitive - they seek to achieve the upper hand or to prevent others from dominating them. For women, however, talking is often a way to gain confirmation and support for their ideas. Men see the world as a place where people try to gain status and keep it. Women see the world as “a network of connections seeking support and consensus”.

B-Independence vs. intimacy

In general, women often think in terms of closeness and support, and struggle to preserve intimacy. Men, concerned with status, tend to focus more on independence. These traits can lead women and men to starkly different views of the same situation

C- Advice vs. understanding

Deborah Tannen claims that, to many men a complaint is a challenge to find a solution:

“When my mother tells my father she doesn't feel well, he invariably offers to take her to the doctor. Invariably, she is disappointed with his reaction. Like many men, he is focused on what he can do, whereas she wants sympathy.” (Tannen 1984:180)

D-Information vs. feelings

Culturally and historically speaking, men's concerns were seen as more important than those of women, but today this situation may be reversed so that the giving of information and brevity of speech are considered of less value than sharing of emotions and elaboration.

E-Orders vs. proposals

It is claimed that women often suggest that people do things in indirect ways - “let's”, “why don't we?” or “wouldn't it be good, if we...?” Men may use, and prefer to hear, a direct imperative.

F- Conflict vs. compromise

This situation can be clearly observed in work-situations where a management decision seems unattractive - men will often resist it vocally, while women may appear to accede, but complain subsequently. In fact, this is a broad generalization - and for every one of Deborah Tannen's oppositions, we will know of men and women who are exceptions to the norm.

7-Views for the reasons of Gender differences between man and woman

This also suggests that explanations of differences between women's and men's speech behaviour which refer only to the status or power dimension are likely to be unsatisfactory. The social distance (or solidarity dimension) is at least as influential. Many of the features which have been identified as characteristic of women's language are positive politeness devices expressing solidarity. And as will be illustrated in the discussion below, there are many other factors which are also relevant when comparing women's and men's use of language, including culture, social role and the formality of the context.

-Examples of Gender differences

Robin Lakoff identified a number of linguistic features which she claimed were used more often by women than by men, and which in her opinion expressed uncertainty and lack of confidence.

-Features of ‘women’s language’

Lakoff suggested that women’s speech was characterised by linguistic features. These features are usually small differences in pronunciation or word-shape (morphology. such as the following:

(a) Lexical hedges or fillers, e.g. you know, sort of, well, you see.

(b) Tag questions, e.g. she’s very nice, isn’t she?

(c) Rising intonation on declaratives, e.g. it’s really good.

(d) ‘Empty’ adjectives, e.g. divine, charming, cute.

e) Precise colour terms, e.g. magenta, aquamarine.

(f) Intensifiers such as just and so, e.g. I like him so much.

(g) ‘Hypercorrect’ grammar, e.g. consistent use of standard verb forms.

(h) ‘Superpolite’ forms, e.g. indirect requests, euphemisms.

(i) Avoidance of strong swear words, e.g. fudge, my goodness.

(j) Emphatic stress, e.g. it was a BRILLIANT performance.

PhD Course in Advanced Sociolinguistics

Language and Identity (Social& Linguistic)

Prepared by Musaab N, Muhammad

Supervised by Asst. Prof. Dr. Wafa’a M, Ali

Introduction

Language is a fundamental aspect of human experience, shaping the way individuals perceive themselves and interact with the world. It is not merely a system of communication but a powerful social tool that constructs and negotiates identity. Through language, people express belonging, differentiate themselves from others, and navigate complex social landscapes. Identity, in turn, is not fixed or inherent; it is fluid, shaped by interactions, ideologies, and power relations.

Linguistic anthropology explores these dynamics, highlighting how language serves as both a medium of inclusion and exclusion, a marker of solidarity and distinction. By examining the ways language reflects and reinforces social structures, this field sheds light on the intricate relationship between communication and identity formation in diverse cultural contexts. Language plays an essential role in shaping both individual identity and broader cultural norms, countering the idea that language merely reflects a person’s identity. This perspective emphasizes the fluid and intricate ways in which social, linguistic, and cultural practices intertwine to define and construct identity.

Identity

The identity is who you are, the way you think about yourself, the way you are viewed by the world and the characteristics that define you. Identity is not something that is inherently fixed but rather the result of cultural processes. It is formed through the interaction of various elements such as practice, ideology, indexicality (the relationship between signs and their meanings), and performance. Ideology refers to the prescribed norms or expectations about how individuals should behave or speak, yet the actual behaviors and actions of people are often more nuanced and strategic.

These behaviors may either challenge or reinforce the ideologies they are influenced by. In other words, identity is a cultural product—it arises from how societies assign meaning to behaviors, language, and social roles. An example of how **identity** is shaped through cultural processes can be seen in the way language is used in different social settings.

Consider a young person who belongs to a multicultural urban community. This individual may be expected by society (based on the ideology of their community) to speak in a certain way—perhaps using a specific dialect or slang associated with their cultural background.

However, in practice, this person may alter their speech depending on the social context. In a professional setting, they might choose to speak in a more formal, standardized manner to align with expectations of professionalism. In contrast, with close friends or in casual environments, they may revert to using more familiar slang or regional expressions.

In this scenario, their identity isn't simply a fixed trait based on their language; rather, it's a dynamic construct influenced by the cultural practices and social expectations they navigate. While the community's ideology suggests a certain way of speaking, the individual strategically performs different identities through their language choices, challenging or confirming the expectations set by society. The interplay between their language, social roles, and the ideologies they encounter helps shape their sense of self and identity within different contexts.

Definitions of Identity

Before going deeper into detail concerning the various issues surrounding identity, we will first take a look at the definition of the concept. When searching for the basic dictionary definition, we find the following one **in the third edition of the Oxford English Dictionary: 'The sameness of a person or thing at all times or in all circumstances; the condition or fact that a person or thing is itself and not something else; individuality, personality'** .When reduced to even more basic wordings, we can quote Paulin Djité who writes that 'identity is the everyday word for people's sense of who they are' (Djité 2006:6**). Identity can be described as the entirety of how we as individuals view or perceive ourselves as unique from others.** Individuals are not free to adopt any identity they like, and factors like their social class, their ethnic group how others see them.

De Fina (2003:15) defines identity within social psychology saying that it is used to refer to the sense of belonging to social categories. Deschamps & Devos (1998: 3) define identity in personal sense by stating that it refers to the fact that each individual is a unique combination of features which makes him different from others and which explains his uniqueness, explaining how an individual is aware of his difference with respect to others since he recognizes himself only as identical to himself, that is, he is the same in time and in space, but that is also what specifies him and marks him out from others. Preece (2016:113) argues that the concept of identity can be defined as the processes by which people affiliate with or distance themselves from specific communities, what kind of information they aim to convey about themselves and how this information in turn resonates with the ideas others hold about them. Individuals have multiple identities, asserting different identities in different circumstances. An individual may, for example define herself primarily as a Muslim in her family or community, as a manager at her work, our names, ages, perhaps addresses reflect some aspects of who we are our birth date and age may suggest certain characteristics or features of our attitudes, clothing styles and music preferences. Our photograph may promote ideas about our ethnic or cultural location. All of these are but superficial markers of an assumed identity Fearson 1999.

The Importance of Identity

I've been thinking a lot lately about what this word means and how it affects our behaviour. We all have a certain image of ourselves - beliefs about the kind of person. Having a strong sense of identity seems to be desirable, something that brings comfort and security. Many people seem to spend a lot of time trying to figure out who they are, what they want, and what they believe. And perhaps rightly so: having a strong identity certainly seems to have advantages. A clear sense of "who you are" makes it easier to connect with other similar people and groups. People with a strong identity often stand out more and are more memorable. I'd guess that the people who are most successful in life - especially people who become famous - all have pretty strong identities.

Identity also helps us to make decisions and to know how to behave. We're constantly faced with complex decisions and circumstances. With no prior beliefs about what we should do, weighing all the options and making a decision would be near impossible. Having a sense of what kind of person you are makes it much easier to decide how you should behave, and to have confidence in your choice between options. This makes decisions that would otherwise be agonising virtually effortless.

But strong identities can also be dangerous. The drive to protect your identity can be overpowering. Sometimes we can get so caught up in this we neglect other important things: like being open-minded, truth-seeking, and kind to others. It's hard to think clearly and objectively about something that you identify strongly with, and I think this is the driving force behind a lot of conflict in the world. Paul Graham has a nice essay in which he advocates for "keeping your identity small" for this very reason - "If people can't think clearly about anything that has become part of their identity, then the best plan is to let as few things into your identity as possible." Another problem with identity is that once established, it can be very hard to change. If you believe that you are a certain kind of person, you'll generally be more likely to (a) act accordingly and (b) interpret your behaviour in accordance with this belief. And of course, this just acts to strengthen your sense of identity further. So, for example, if you think of yourself as an altruistic person, you're probably more likely to seek out opportunities to help people, and to interpret your behaviour as altruistically driven. Whereas if you think of yourself as a lazy person, you're probably more likely to spend whole days watching TV in bed, and to think of times that you struggle with work as being down to laziness.

This self-reinforcing nature of identity is a double-edged sword. For positive aspects of your identity, it can be helpful - if you believe that you're altruistic, this belief is likely to make you more altruistic. But if you have parts of your identity that aren't so rosy (and don't we all?) this kind of circularity can be quite damaging, because it makes identity very difficult to change. It's hard for the person who thinks they are lazy to change this belief, because to do so he needs to change his behaviour - which is difficult to do if you identify as lazy. This loop seems pretty tough to break out of. So what should we do about this? Simply making your identity as small as possible doesn't seem to be the solution - what you really want to do is only keep those parts of your identity that are helpful, and discard any that hold you back.

Can we clearly distinguish between "helpful" and "unhelpful" identities? Probably not. There are some clear cases - having as part of your identity things like "being open-minded", "having a growth mindset", and "being a kind and considerate person" seem to be pretty straightforwardly useful. Identifying as someone lazy or boring or stupid seem pretty uncontroversially harmful.

But there's a lot of middle ground - especially when it comes to identities that are tied to certain groups or ideologies - identifying as right-wing or left-wing, utilitarian or deontologist, religious or atheist. Whilst these kinds of identities can be good for connecting with those similar and help with success, they can also polarize groups and cause conflict.

One solution here would be to stop using identity labels, and instead just talk about beliefs - instead of asking someone "Are you Tory/Lib Dem/Labour?", we could simply ask "Do you agree with policy x?". I personally feel quite uncomfortable labelling myself as "a" anything. I tend to shy away from these kind of identity questions when asked, and focus more on what I believe. The approach of focusing more on beliefs seems likely to be more time consuming. It's much easier to simply attach a rough label to someone than to discern all of their different beliefs. Perhaps this is ultimately what it comes down to - identity labels make it much easier for us to model both ourselves and other people. And whilst we can try to be careful with the identities we attach to ourselves and others, it's near impossible to abandon them altogether.

Types of Identity

1- Social identity

Social identity refers to the way individuals define themselves in relation to the groups they belong to. It encompasses the various social categories that people use to define who they are, such as nationality, ethnicity, religion, gender, sexual orientation, social class, occupation, or any other group membership. Social identity plays a significant role in shaping an individual's behavior, experiences, and perceptions of themselves and others.

A social identity is a person knowledge's that he or she belongs to a social category or group. A social group is a set of individuals who hold a common social identification or view themselves as a members of the same social category.

A key theory of group identity is social identity theory. Social identity theory developed from a series of studies, conducted by the British social psychologist Henri Tajfel and his colleagues in the early 1970s. Tajfel (1979) proposed that the groups (e.g social class, family, football team…etc.) which people belonged to were an important source of pride and self-esteem. Groups give us a sense of social identity: a sense of belonging to the social world. In order to increase our self-image we enhance the status of the group to which we belong. Therefore, we divided the world into "them" and "us" based through a process of social categorization (i.e. we put people into social groups). This is known as ingroup (us) and out-group (them). Social identity theory states that the in group will discriminate against the out-group to enhance their self-image. Henri Tajfel proposed that stereotype (i.e. putting people into groups and categories) is based on a normal cognitive process: in doing so we tend to exaggerate.

1. The differences between groups
2. The similarities of things in the same group.

We categorize people in the same way. We see the group to which we belong (the in-group) as being different from the others (the out- group), and members of the same group as being more similar than they are. Tajfel and Turner (1979) proposed that there are three mental processes involved in evaluating others as "us" or "them" (i.e. "in-group " and "outgroup "). **These take place in a particular order**.

The first is **categorization** we categorize objects in order to understand them and identify them. In a very similar way we categorize people according to their interest ethnicity, gender, occupation and other factors. If we can assign people to a category then that tells us things about those people, and as we saw with calling someone Muslim, footballer, mother, students…etc. As ways in which we do this. We find out things about ourselves by knowing what categories we belong to. We define appropriate behavior by reference to the norms of groups we belong to. An individual can belong to many different groups.

In the second stage, **social identification**, we adopt the identity of the group we have categorized ourselves as belonging to. If for example you have categorized yourself as a student, the chances are: you will adopt the identity of a student and begin to act in the ways you believe students act(and conform to the norms of the group), and your self-esteem will become bound up with group membership. **The final stage is social comparison**. Once we have categorized ourselves as part of a group and have identified with that group we then tend to compare that group with other groups. If our self-esteem is to be maintained our groups needs to compare favorably with other groups.

Basic Concepts of Social Identity:

A-Group Membership: At the core of social identity is the idea that individuals identify with specific social groups. These groups provide a sense of belonging and shared experiences. Group membership could be based on things like:

Ethnic or Racial Groups: Common cultural background, language, and traditions. It is what makes individuals unique and different from others and it is self-defined. It denotes how strongly an individual is attached, relates, or feels a sense of belonging to a particular ethnic group or culture. It signifies a sense of continuity of belonging to a group that shares certain elements such as cultural practices, language, interactions with society and ethnic peers, as well as preferences for same ethnic friends or colleagues. Drydakis(2012), using data from Greece, suggested that ethnic identification is a combination of language, cultural habits (food, media, music and reading), self-identification, social interaction, and future citizenship plans .

Religious Groups: can be a part of how you identify yourself as a person. It is identity that is based on your **inclusion in a particular faith** and the sense. Religious consumption norms can serve as communicative devices that lower the cost of expressing one's religious identity. Each religion typically provides a distinct set of consumption norms that become the blueprint for its followers. For example,

- Muslim women wear headscarves to cover their hair
- Jewish men similarly wear Yarmulke for religious expression other examples of religious expression include distinctive styles in clothing and grooming, necklaces with the cross, household decorations, and items like candles and ornaments that mark religious holiday.

Gender and Sexual Groups: Identification with a particular gender or sexual orientation. Gender identity is defined as a personal conception of oneself as Male or female (or rarely, both or neither), gender identity is manifested within society by observable factors such as behavior and appearance. For example, if a person considers himself a male and is most comfortable referring to his personal gender in masculine terms, then his gender identity is male. However, his gender role is male only if he demonstrates typically male characteristics in behavior, dress, and so on.

Geographical or National Groups: Affiliation with a particular country, region, or community. It is the notion that part of a person's identity is rooted not only in the country but also in the region they live in. A strong example of regional identity would be if one was asked where he or she was from and would reply " Mosul" instead of Iraq. Regional identity, then, is a sense of belonging similar to that of national identity but on a smaller scale or level.

National Groups: is an ethical and philosophical concept where by all humans are divided into groups called nations, members of a "nation " share a common identity, and usually a common origin, in the sense of ancestry, parentage or descent. The national identity represents the socio-historical context within which culture is embedded and the means by which culture is produced, transmitted and received "(Guibernan, 1996:79) national identity expresses singular and fixed instead of multiple and dynamic. **National identity is based upon the sentiment of belonging to a specific nation**, endowed with its own symbols, traditions, sacred places, ceremonies, heroes, culture and territory.

Personal Groups: How a person defines who he or she is , self-definition or self understanding simply, it refers to how a person thinks about themselves. Is the concept you develop about yourself that evolves over the course of your life. This may include aspects of your life that you have no control over, such as where you grew up, or the color of your skin, as well as choices you make in life such as how you spend your time? And what you believe.

Master Groups: Master (primary) identities are relatively stable and unchanging: gender, ethnicity, age, national and regional origins. The meanings of master identities change across time and space. Master identity influence the production of talk and how interlocutors alter cast speaker on the basis of language that they hear. It is the interplay between speakers production and interlocutors alter casting of identity by shifting language style and by alternating language code.

Intersectional Groups: International identities refer to roles that people take on in a communicative context with specific other people. Interactional identity are inherently variable because of the diversity of role that individuals play in a community. The same person can be a teacher, a student, a mother, a partner in interactions with different people and in different physical surrounding and the way of speaking that differ according to the identity role. Teachers conversations with students in a classroom often take of the well-known three part sequence of initiation by, response by student, and evaluation or feedback by teacher.

Cultural Groups: is the identity or feeling of belonging to a particular religion, social class, locality or any kind of social group that has its own culture. Cultural identity is constructed and maintained through the process of sharing collective knowledge such as traditions, heritage, language, norms and customs.

Political Groups: In this type of identity, in fact, the focus would be on the identity of political party which can basically be defined as a group of people who come together to share the same political ideas, beliefs, values, and behavioral orientations and ideology, including a sufficient degree of identification with that specific political identity (Wiarda, 2014: 1). Moreover, political identity is not a purely emotional or even irrational thing; it rather contains reflections and emotions, thoughts and images, beliefs and symbols. It comes up and is communicated among citizens, groups and generations with the help of symbols, which are by the way not limited to official para-national items such as the flag, anthem (Lucarelli et al, 2011 :4-5).

B- In-group vs. Out-group: Social identity often involves the differentiation between "in-groups" (groups with which an individual identifies) and "out-groups" (groups to which an individual does not belong). This distinction can impact how people perceive others, leading to: In-group favoritism: Tendency to favor members of one's own group over others.

Out-group bias: Tendency to stereotype or hold negative views about those not in one's group.

2-Linguistic identity

The linguistic identity is part and often an important part of our identity; it is basing one's own identity on the language they commonly use, especially their native language. It refers to a person's identification as a speaker of one or more languages. Linguistic identity may refer to the sense of belonging to a community as resource of language, or to the varying ways in which we come to understand the relationship between our language and ourselves. Linguistic identities become central in globalization due to the fact that movement of people, ideas, products, and cultural forms across national boundaries intensifies contact among languages. The role of language in the development of identity in particular is intrinsic to the expression of culture. Language is a fundamental aspect of cultural identity. It is the means by which we convey our innermost self from generation to generation. It is through language that we transmit and express our culture and it is values. Language has the power to define and shape the human experience, it is because of language that I can name my experiences.

1.The Role of Language in Identity Construction

Language is central to identity formation because it serves as a primary means of communication and self-expression. People often identify with specific languages, dialects, or speech communities, and these choices reflect personal, social, and cultural values. For instance, the decision to speak a particular language or dialect can express solidarity with a particular group, region, or nation. It can also signify membership in a social class or a professional group.

Linguistic identity is not static; it evolves as individuals and communities move through different contexts. An individual might speak one language at home, another in the workplace, and yet another in public spaces. These shifts highlight how language choices are intricately tied to one's identity in a specific setting, shaping how individuals present themselves and how they are perceived by others.

2. Dialect and Accent as Markers of Identity

Dialect and accent are crucial components of linguistic identity. They are not just about the words people use but also how they pronounce them. For example, someone's regional accent or dialect can signal where they are from, their social background, or even their cultural affiliation. In some contexts, accents can carry social weight—certain accents may be viewed as prestigious or authoritative, while others may be seen as less "refined."

However, linguistic identity based on dialect or accent can also be a source of discrimination and stigmatization. People who speak with non-standard accents or use regional dialects may be marginalized or stereotyped based on perceived associations with lower social status or lack of education. This illustrates the intersection of linguistic identity with power dynamics and social hierarchies.

3. Multilingualism and Fluid Identities

In the globalized world, multilingualism is becoming more common. Individuals who speak multiple languages navigate complex identities that incorporate different cultural and linguistic frameworks. In such cases, linguistic identity can become fluid and adaptable, changing according to the social, cultural, or professional context. A person might identify as a "native" speaker of one language in one community but may feel a strong affinity for another language in a different context. Multilingual speakers often use code-switching—the practice of alternating between languages or dialects within conversations—as a tool for expressing different aspects of their identity, whether they're drawing on a specific cultural experience, social group, or professional background.

While multilingualism can be a source of pride, it can also present challenges related to identity. For example, individuals might feel torn between the cultures and languages they speak, leading to a sense of fragmentation or conflict within their linguistic identity. This can be especially evident among immigrant populations, whose linguistic identity might be shaped by both the language of their heritage and the language of their adopted country.

4. Social and Political Dimensions of Linguistic Identity Linguistic identity also has a social and political dimension, particularly when it comes to issues of language rights, language preservation, and language policy. In many regions, the dominance of a particular language over others—whether due to historical, political, or economic factors—can have profound implications for linguistic identity. For example, speakers of minority languages often find their linguistic identities threatened by the dominance of a global or national language.

Language can be a powerful tool for resistance or empowerment. For example, indigenous groups and linguistic minorities often seek to reclaim their languages as part of broader movements for cultural autonomy and recognition. In such cases, linguistic identity becomes a form of resistance against cultural erasure and a way of asserting political and social rights.

Views towards Social and Linguistic Identity

1- Functionalist view of identity

Functionalism is a theory about the nature of mental states. According to functionalism, mental states are identified by what they do rather than by what they made of. The main arguments for factionalism depend on showing that it is superior to its primary competitors: identity theory and behaviorism. Contrasted with behaviorism, functionalism retains the traditional idea that mental states are internal states of thinking creatures. Contrasted with identity theory, functionalism introduces the idea that mental states are multiply realized. Functionalism is the theory that mental states are more like mouse traps than they are like diamonds. That is, what makes something a mental state is more a matter of what it does, not what it is made of. This distinguishes functionalism from contemporary monisms such as J. J. C. Smart’s mind brain identity theory. The identity theory says that mental states are particular kinds of biological states—namely, states of brains—and so presumably have to be made of certain kinds of stuff, namely, brain stuff. Mental states, according to the identity theory, are more like diamonds than like mouse traps. Functionalism is also distinguished from B. F. Skinner’s behaviorism because it accepts the reality of internal mental states, rather than simply attributing psychological states to the whole organism.

2-**Marxist view of identity**

Marxists argue that the most significant part of a person's identity is their social class. A person's social class determines the way that a person sees the world around them. For Marxists identities are shaped by the classes conflicting with each other. (Theories such as functionalism and Marxism see identity as originating in fairly straightforward way from involvement in particular culture, and subcultures, e.g. people living in Britain would be expected to have a strong sense of British identity) .

3-Interactive perspective view of identity

Interactionism is a theoretical perspective that derives social processes (such as conflict, cooperation, identity formation) from human interaction. It is the study of how individuals shape society and are shaped by society through meaning that arises in interactions. In The Presentation of Self in Everyday Life, Goffman describes the techniques used by the individual to create and maintain a certain impression. In this context he utilizes terms borrowed from the theatre, such as interaction (mutual influence of individuals on each other's actions), presentation (activities of an individual to influence others), or role (the pattern of action being performed, including the standardized repertoire of expressions used in a situation).

4- **Labeling view**

Labeling view stems from the symbolic interactionism tradition of the " looking -glass self" (Cooley 1902) or the reflexive self (Mead 1934), which means that an individual will come to view the self in a way that reflects the views of others and thus come to act in a manner consistent with these views. According to labeling theory, this label becomes internalized and forms part of the identity of the person, who subsequently comes to act out on the basis of this identity. Labeling theory suggests simply that an individual will behave in accordance with a label applied to him or her by others. Labeling theory suggests that labeling occurs when the behavior of an individual is identified by another person as deviant or wrong (Becker, 1963; Hagan, 1973). It connotes a reciprocal relation between the actions of an individual and society's response to that act.

Conclusion

The conclusion emphasizes the importance of continued exploration of identity within the field of linguistic anthropology. It highlights that although identity has always been a part of language and culture studies, it has only more recently been explicitly examined. The authors call for a more defined theoretical approach to understanding identity, suggesting the integration of ideas like intersubjectivity, semiotics, and power relations. They propose a model of identity that takes into account aspects like markedness, essentialism, and institutional influence. Additionally, they acknowledge the critiques of both the field and the concept of identity itself, urging these concerns to be addressed in future research.

References

De Fina, A Schiffrin, D. &M Bamberg (Ed.) (2003). Identity in Narrative: A Study of Immigrant Discourse. Amsterdam: Benjamins Publishing Company.

Djité, Paulin G. 2006. Shifts in linguistic identities in a global world. Language Problems & Language Planning 30(1).1-20.

Edwards, J (2009). Language and Identity. Cambridge: Cabridge University Press.

Florian, C. (1998). The Handbook of Sociolinguistics. Blackwell Publishing.

Tajfel, H. (1978). Differentiation Between Social Groups: Studies in Intergroup Relations. London, England: Academic Press.

Wiarda, H, J. (2014). Political Culture, Political Science, and Identity Politics: An Uneasy Alliance. England: Ashgate Publishing.

Sociolinguistics / 2025

Linguistic Imperialism

PhD Candidate Ahmed Abdulrazzak Aziz

Supervised by: Asst. Proff. Dr. Wafa’a Mudhaffar Ali

Historical Background

The term linguistic imperialism **(LI)** was first introduced in scholarly circles by Phillipson. He initially used the term to refer exclusively to the global status and role of English and applied the term primarily to English language teaching. For Phillipson, LI is a situation in which "the dominance of English is asserted and maintained by the establishment and continuous reconstitution of structural and cultural inequalities between English and other Languages".

The notion of LI is not restricted to English. LI can be explored in relation to other languages in other historical periods and regions. Latin hegemony in the Roman Empire

around the first to the fourth century, Arabic hegemony over Europe around the eighth to the thirteenth century, and Japanese hegemony in the East Asian region in late nineteenth and early twentieth century are examples of LI by other languages. However, there is a good case to be made that LI by English is different in degree and kind.

Though English was first spread in many countries in South Asia, Africa, and the Caribbean by the British empire, English received a new lease on life after decolonization when the United States gained ascendance after the Second World War. The types of economic, cultural, and political power the USA wields in the world today further the currency and status of English. In some ways, the power of English is not dependent on a specific country anymore. It is sustained by transnational processes and institutions.

Defining the linguistic imperialism

Linguistic imperialism (LI) refers to the imposition of a language on other languages and communities. It is an exercise of power at the transnational level with geopolitical implications.

Linguistic imperialism is the imposition of one language on speakers of other languages. It is also known as linguistic nationalism and linguistic dominance. In our time, the global expansion of English has often been cited as the primary example of linguistic imperialism.

Linguistic imperialism is occasionally defined as the transfer of a dominant language to other people. This language "transfer" comes about because of imperialism. The transfer is considered to be a demonstration of power; traditionally military power but also, in the modern world, economic power.

Linguistic imperialism is a form of linguicism which benefits and grants power to the dominating/oppressing language and its speakers.

Central Constructs

Ideologies are ideas, assumptions, attitudes, and values that explain the unequal status of

individuals and communities in society.

Discourses are ideologies that have found expression in language. Discourses are genres

of thinking and communicating that have social and political functions.

Hegemony is exercised when the ideologies and discourses of a powerful community are internalized by other social groups, to the extent that they willingly participate in the leadership of that community. For example, the discourse that English is a superior language with the capacity to express complex philosophical, scientific, and technical information gains ideological implications when multilingual communities believe this discourse and learn English at the cost of proficiency in their own languages. When they internalize this discourse and buy into the forms of knowledge, values, and identities that come with English, they become unwitting participants in the power enjoyed by English and Anglophone countries.

There are different **labels** for distinguishing between countries that have unequal relationships based on their language identity. Phillipson (1992) used the terms **center** and **periphery**. The terminology has the advantage of connecting language inequality to other forms of economic and political inequalities. Kachru (1986) introduced the terms **expanding circle** (where English is used as a foreign language for contact purposes with outsiders – i.e. Vietnam or Angola, which were not former British colonies), **outer circle** (where English is a second language with its own well established varieties since colonial times – i.e. India, Nigeria), and the **inner circle** (where ownership of English and native speakerhood have been traditionally claimed – i.e. the UK, the USA, Canada, Australia). Kachru classified these communities as **norm-dependent**, **normdeveloping**, and **normproviding**, respectively, to indicate their relative status in relation to grammatical standards.

Linguicism

Linguicism refers to “ideologies, structures, and practices which are used to legitimate, effectuate, and reproduce an unequal division of power and resources (both material and immaterial) between groups which are defined on the basis of language”.

The term is analogous to racism and sexism, and refers to a discriminatory attitude towards language that is played out in social practices and sustained by social institutions. For example, the lack of policy emphasis and funding for teaching languages other than English in the USA can be considered an institutionalized form of linguicism.

Linguicism leads to the promotion of certain languages and language varieties and the stigmatization of others, as the prestigious language becomes the norm by which other languages derive their status.

The implications of linguicism for LI – and for the LI of English in particular – is that English is equated with prestige, while failure to use English or even using other languages may connote lack of status. Oda (2000) demonstrates that those who are not proficient in English are treated as lower in status in Japanese universities.

Native speakerism

The discourse of native speakerism is made of the following assumptions:

1. that “native” speakers are the authorities on the language and enjoy superior competence;
2. that those who use it as an additional language have to treat “native” speaker competence as the target;
3. and that “native” speakers are the best qualified to teach that language.

Counter evidence to native speakerism

Other scholars argue against native speakerism:

1. there are no status differences between languages in purely linguistic terms,

2. language change or diversification cannot be stopped by attempts at purification,

3. language learning is a creative cognitive and social process that has its own trajectory, and is not fully dependent on the teacher.

Dilemma

Many scholars consider the term ‘native’ speaker as itself questionable. Some in postcolonial communities acquire English simultaneously with one or more local languages to develop multilingual competence. These speakers would consider themselves first language speakers of English as well as of one or more local languages.

We should treat multilinguals as endowed with a **multicompetence** that is qualitatively different from the competence of monolingual speakers of English.

Implication

Such developments call for **new ways of classifying language identities**. Rampton (1990) has argued for categorizing linguistic identities not in terms of birth, but in terms of a more diverse set of categories such as **expertise**, **affiliation**, and **inheritance**.

Monolingualism

Another discourse that facilitates LI is monolingualism. This discourse has implications for many subfields in applied linguistics.

1. in second language acquisition, processes of learning treat monolingual acquisition as the model,
2. in language planning, multilingualism is treated as a problem for social progress,
3. in sociolinguistics, identities and communities are defined in terms of homogeneity,
4. in teaching, English is best taught monolingually

In teaching. This discourse takes as its model the pedagogical norms used in countries where English is the primary language, and assumes that speaking English to the exclusion of other languages (including the mother tongue) is the best way to optimize the learning of this language in second language contexts. Example, “English only” is institutionalized in many states of the USA, such as Arizona, where other languages are not permitted to be used in classrooms.

In colonized countries, local people may themselves desire to be taught in the

English language in order to reach a better economic status or because this language represents a valued cultural and economic capital.

Counter evidence to monolingualism

There are several factors that make the monolingualist discourse a fallacy:

1. when the mother tongue is banned from the classroom, the teaching leads to the **alienation of the learners**, deprives them of their cultural identity, and leads to acculturation rather than increased intercultural communicative competence,

2. the use of first language (L1) can help students **bridge** home knowledge and school knowledge more effectively,

3. the provision of mother tongue in the educational process (including the L2 classes) is a fundamental "**linguistic human right'** of the minority language groups,

4. the use of students' first language can **increase their openness** to learning English

by reducing the degree of language and culture shock.

Defining characteristics of linguistic imperialism

Phillipson argues that the defining characteristics of linguistic imperialism are:

1. As a form of linguicism, which manifests in favoring the dominant language over another along similar lines as racism and sexism.

2. As a structurally manifested idea, where more resources and infrastructure are given to the dominant language.

3. As being ideological, in that it encourages beliefs that the dominant language form is more prestigious than others. These ideas are hegemonic and internalized and naturalized as being "normal".

4. As intertwined with the same structure as imperialism in culture, education, media, and politics.

5. As having an exploitative essence, which causes injustice and inequality between those who use the dominant language and those who do not.

6. As having a subtractive influence on other languages, in that learning the dominant anguage is at the expense of others.

7. As being contested and resisted, because of these factors.

According to Robert Phillipson (2013), Linguistic imperialism, operates within broader structures of imperialism encompassing culture, education, media, communication, economy, politics, and military activities. It systematically privileges the dominant language, reinforcing hierarchy, inequality, and exploitation through structural and ideological mechanisms.

Structural and Ideological Dimensions

Hegemonic Dominance: The dominant language is internalized as "normal," reinforcing linguistic hierarchy.

Linguicism: A linguistic parallel to racism, sexism, and classism, privileging users of the dominant language.

Subtractive Language Learning: Acquisition of the dominant language occurs at the expense of local languages, leading to linguistic displacement.

Glorification and Stigmatization: The dominant language is framed as superior (e.g., logical, modern, civilizational), while subordinated languages are diminished as mere dialects or obstacles to progress.

Historical Context and Mechanisms of Expansion

The historical trajectory of linguistic imperialism is evident in colonial language policies: European Colonialism: European languages were transplanted globally through military conquest, commerce, and missionary activity, often marginalizing indigenous languages (Fanon, 1952; Mühlhäusler, 1996).

Education as a Tool of Imperialism: Colonial education systems facilitated linguistic hegemony, as seen in the British and French empires. English-medium education, for instance, created a socio-economic elite while restricting linguistic rights for the majority (Ngũgĩ, 1985; Mazrui, 1997).

Postcolonial Continuities: The dominance of English persists post-independence, deepening socio-economic stratification, particularly in India and Africa (Mohanty, 2006). Institutions like the World Bank reinforce this through funding policies that privilege former colonial languages over indigenous languages.

Contemporary Manifestations

Neoliberal Globalization: The expansion of English aligns with US and UK geopolitical and economic interests, facilitated by global institutions and corporate influence (Piller & Cho, 2013).

English as a Lingua Nullius: The marketing of English as a neutral global lingua franca obscures its role in reinforcing global inequalities (Phillipson, 2017).

Military and Economic Imperialism: NATO expansion and US military interventions consolidate the linguistic dominance of English (Templer, 2016).

Education and Linguistic Elitism: The proliferation of English-medium international schools creates a transnational elite detached from their local linguistic and cultural roots (Wechsler, 2017).

Resistance and Counteraction

Policy Interventions: Nordic countries implement bilingual education policies to balance English proficiency with the preservation of national languages. The EU faces pressures of English dominance but maintains multilingualism through institutional policies.

Linguistic Rights and Activism: Suppression of linguistic minorities (e.g., Kurds in Turkey, Uyghurs and Tibetans in China) exemplifies extreme cases of linguistic imperialism. Theoretical frameworks and empirical studies (Skutnabb-Kangas & Phillipson, 2017) highlight the need for linguistic justice.

References

- Phillipson, R. (1992). Linguistic Imperialism. Oxford University Press.

- Phillipson, R. (2013). Linguistic Imperialism. *The Encyclopedia of Applied Linguistics.* Blackwell publishing Ltd.

- https://en.wikipedia.org/wiki/Linguistic_imperialism

- https://www.thoughtco.com/what-is-linguistic-imperialism-1691126#:~:text=Linguistic%20imperialism%20is%20the%20imposition,primary%20ex

ample%20of%20linguistic%20imperialism.

Sociolinguistics

Corpus linguistics

&

Doing Discourse Analysis in Sociolinguistics

PhD Candidate Ahmed Abdulrazzak Aziz

Supervised by: Asst. Proff. Dr. Wafa'a Mudhaffar Ali

Corpus linguistics

Introduction

Corpus linguistics employs large, electronically stored collections of naturally occurring language data, analyzed through specialized software to investigate linguistic patterns. Methodological techniques such as keyword analysis, frequency lists, and concordance analysis enable researchers to explore sociolinguistic variation systematically.

Corpus construction involves selecting representative texts, which can be used to examine frequency patterns, lexical bundles, and tagged grammatical features. A corpus-driven approach allows for the identification of unexpected linguistic differences and similarities between social groups, as opposed to hypothesis-driven methods.

By comparing corpora—for example, speech data from younger versus older adults—researchers can uncover statistically significant keywords and collocational patterns that reveal social and stylistic variation. This approach supports empirical, large-scale analysis and reduces the potential for introspective bias, though it also requires careful attention to corpus design and contextual interpretation.

Corpus linguistics

Corpus linguistics is the study of language using a corpus, which means a large collection of texts stored in electronic form. These texts represent a specific language or variety and are studied using special software tools. A basic analysis includes making a frequency list—a list of how often each word appears. We can also study n-grams (word combinations like "I'm very sorry") and different parts of speech (like nouns or adjectives), which requires tagging—labeling each word with extra information like its grammatical role.

Researchers also look at dispersion, which checks if a word appears in many texts or just one. When comparing different corpora, we can find keywords—words that appear significantly more in one corpus than another. These help us discover important patterns in language use.

Another important concept is collocation, which is when two words often appear near each other, like “elderly” and “infirm.” This shows how words can be connected in meaning. We can also do concordance analysis, which shows each use of a word in context. This helps us understand how the word is used in real sentences and involves qualitative analysis—reading and interpreting the data carefully.

The advantage of using a corpus is that it gives us evidence-based findings based on real and large-scale language use. This avoids errors that can happen when relying only on introspection (thinking about language without data) or small data samples, which can lead to biased or cherry-picked results.

Building a Corpus

Building a corpus begins with defining the research questions and setting constraints for data collection. For instance, if the question is "Are women politer than men?", the focus would be on collecting spoken data from both men and women across various contexts and backgrounds. This ensures representativeness, meaning the corpus should reflect the broader population.

A main consideration is balance: for example, ensuring an equal number of male and female participants and representative samples from different settings (e.g., public and private conversations). Care must be taken to avoid sampling bias, such as selecting only the beginnings of texts. Ideally, samples should be taken from different parts of the texts (e.g., beginnings, middles, ends).

Once the data is collected, it needs to be digitized and stored electronically. Written texts are typically easier to digitize, while spoken data require recording and transcription, often with the inclusion of coding schemes for features like pauses, laughter, or volume. After this, tagging is used to provide additional metadata about each text, such as the author's sex, date, or file information, which helps in organizing and analyzing the data. In the case of spoken corpora, speakers' attributes (e.g., age, social class) can also be tagged, enabling detailed comparisons (e.g., analyzing speech patterns by female speakers under 20).

While tagging is not always essential for small, personal projects, it becomes important for larger, shared corpora, especially when conducting complex analyses of grammatical or semantic features. Proper tagging ensures the corpus can be used for diverse analyses by different researchers.

Research Questions

Corpus linguists often formulate research questions that guide their analysis. One approach is corpus-based research, where the researcher starts with a specific theory or hypothesis and identifies what to look for in advance, such as analyzing how age influences apologies. The corpus serves as the foundation for testing the theory. Alternatively, corpus-driven research is more exploratory, where the researcher may not have a specific hypothesis in mind but is guided by the data, such as discovering keywords or patterns through frequency analysis, which leads to new lines of investigation (e.g., "What distinguishes this corpus?" or "How do male and female languages differ?").

In practice, many studies combine both approaches. For instance, a study on age and apologies may start with a list of known apology types but may uncover new ones during analysis, thereby shifting the research direction. Sociolinguists often focus on questions related to specific linguistic features (e.g., how people swear) or social groups (e.g., how women swear). They may also investigate independent variables (e.g., gender) and examine differences (e.g., "Do women swear more than men?") or similarities between social groups, recognizing that linguistic similarities may outweigh differences.

Research questions can be descriptive or involve more detailed analysis of context (e.g., "In what contexts do men and women use feature X differently?") or explanation (e.g., "Why do men and women use feature X differently?"). These questions can be explored through tools like concordance lines, though sometimes additional external information may be needed to fully answer the questions.

Comparing the Speech of Younger and Older Adults

Summary using text terminology in clear academic language:

This study uses the British National Corpus (BNC) to compare the speech of two age groups: adults aged 25–34 and older adults aged over 60. The aim is to investigate lexical differences between these groups through a keyword analysis. The data from these two groups—1,120,516 words from the younger adults and 1,137,433 words from the older adults—were compared to identify statistically significant keywords. A cut-off significance level of 0.01% was applied, focusing on the 50 strongest keywords.

The results reveal several lexical differences between the two groups, which were categorized primarily by grammatical features. For example, war appeared significantly more often in the older group’s speech, likely due to the context of oral history interviews. This context-specific usage was further confirmed by concordance analysis, revealing that older speakers were prompted to discuss the war by interviewers, rather than bringing it up spontaneously in casual conversation. Similarly, the older group showed a greater frequency of past tense forms (e.g., was, were), while the younger group used more present tense forms (e.g., is, are), reflecting their different conversational contexts.

Other notable differences include the use of "daddy" by younger adults, indicating their role as parents. Additionally, discourse markers varied: younger adults favored yeah, okay, and right, while older adults used mm, yes, and the dialectical aye. The younger group also exhibited a higher frequency of the hesitation marker "erm", compared to "er" in the speech of older adults.

While the keyword analysis highlights lexical differences, it is less useful for identifying similarities. Therefore, a frequency analysis of the most common words in both groups was conducted, showing that the two groups shared many common words, with 20 of the top 25 words being identical. These findings demonstrate that despite lexical differences, there are notable linguistic similarities between the two groups.

This study highlights the utility of corpus linguistics in sociolinguistic research, particularly in uncovering unexpected findings, such as context-dependent lexical usage. It also emphasizes the importance of contextual analysis via concordance lines to avoid misleading conclusions, as seen with the impact of oral history interviews in the older group’s speech. Further research could refine these results by excluding interview data to assess whether the observed differences persist.

Quagmires and Troubleshooting

- **Complexity of Language Features**: Corpus linguistics methods may not always be suitable for studying complex or variable language features. For instance, analyzing how males and females use metaphors or tell stories is challenging because it is difficult to define precise search terms for these abstract features. Even when trying to identify conflict, simple word searches (e.g., “no,” “disagree”) may return irrelevant results or miss important instances.

Some researchers have attempted manual methods, such as reading through the entire corpus, which is time-consuming and subjective.

- **Accessing and Transcribing Data**: Written corpora are easier to compile due to the availability of online archived texts like books and articles. However, spoken corpora are harder to collect and transcribe, raising ethical concerns related to consent and anonymity. Spoken corpora are typically smaller than written ones, limiting the ability to generalize findings. Additionally, even large spoken corpora, such as the Corpus of Contemporary American English (COCA), may not accurately represent private, unscripted speech, as they primarily include public media speech. Thus, researchers must be cautious in drawing conclusions about natural conversation from such sources.

- **Overgeneralization**: Researchers must avoid overgeneralizing based on a single linguistic distinction. For example, in the British National Corpus (BNC), females use the word "lovely" more frequently than males. However, a closer examination reveals that this pattern is influenced by age, as older females use "lovely" more than younger ones. Thus, conclusions like "females use lovely more than males" oversimplify the situation and fail to account for the nuanced interactions of multiple social factors. Dividing speakers into more specific subgroups (e.g., males aged 25–34 from social class AB in northern England) can further reduce the sample size and make generalizations more difficult.

- **Feature Distribution Within Social Groups**: The distribution of linguistic features across a social group can influence interpretation. A feature might appear frequent due to a small number of speakers using it heavily, as seen in Harrington (2008), where a few women reported others' speech significantly more often than men, skewing the overall results. Therefore, examining how a feature is distributed across speakers within a group is crucial to understanding the true pattern.

- **Contextual Interpretation**: Corpus data should not be interpreted in isolation, as context is essential for accurate analysis. For example, the word "hi" may appear more frequently in younger speakers' speech in the BNC, but a simple frequency count overlooks its various uses (e.g., as a false start or part of another word). A detailed concordance analysis reveals that many instances of "hi" are not greetings, but rather part of other expressions, demonstrating the importance of context in interpreting corpus data.

Doing Discourse Analysis in Sociolinguistics

Introduction

Discourse analysis is a versatile method employed across various disciplines, including anthropology, sociology, philosophy, and social psychology, to explore the social and structural dimensions of language. Linguists use discourse analysis to study how texts are organized and constructed, while sociolinguists focus on the role of language in social interaction, examining how language reflects and shapes social relationships.

Discourse analysis addresses a wide range of research questions, such as how social roles like friendship are enacted through language, how power and leadership are constructed, and how individuals manage conversational turn-taking. It also explores gendered language use, for instance, by examining whether women or men use more hedges in conversation, and why such differences may occur. Furthermore, discourse analysis is used to investigate the structure and cultural context of formal interactions like interviews and meetings, shedding light on cross-cultural variations in communicative practices.

Sociolinguists have applied discourse analysis to study how people give compliments, make complaints, engage in humor, and manage the openings and closings of conversations and meetings. The method is also employed to understand the dynamics of miscommunication in interactions.

In essence, discourse analysis is a valuable tool in sociolinguistics for examining the social meanings conveyed through language in interaction, providing insight into how language functions within specific social contexts. This chapter outlines how sociolinguists conduct discourse analysis and highlights its utility in revealing the social significance of language use.

Theoretical frameworks

Discourse analysis can be employed within various theoretical frameworks, each with distinct goals and methods for analyzing language in use.

- **Critical Discourse Analysis (CDA):** This approach focuses on identifying how power dynamics are enacted through discourse, often in covert ways. CDA examines how seemingly neutral discourse structures (e.g., speaking rights, agenda-setting, or passive voice) can favor those in power and marginalize the powerless. Researchers using CDA are eclectic in their methodology, drawing from various approaches to analyze discourse and its social implications (van Dijk, 2001; Wodak, 2001).

- **Variationist Sociolinguistics:** This framework, often associated with social dialectology, aims to quantify specific discourse features (e.g., tag questions or discourse markers like "you know") and correlate them with social categories such as gender, social group, or ethnicity (Tagliamonte, 2006, 2012; Cheshire, 2007). Variationist sociolinguists track changes in language use across social groups and over time, studying the spread of linguistic features like the New Zealand tag "eh" or the quotative "I'm like." Challenges include identifying the full range of meanings for these features and determining appropriate positions for quantification (Pichler, 2010; Tagliamonte, 2006).

- **Conversation Analysis (CA):** Rooted in sociology and ethnomethodology, CA focuses on the sequential organization of discourse and how speakers interpret each other's utterances within a conversation. Researchers look for patterns across multiple instances of talk, paying attention to micro-level features like pauses, hesitations, and volume changes. CA emphasizes the importance of detailed transcripts to capture the fine-grained details of interaction and avoids making inferences about the wider social context unless explicitly indicated by the participants themselves.

- **Interactional Sociolinguistics:** This approach emphasizes the social context of discourse and aims to identify how various social factors (e.g., age, ethnicity, or stances like aggression or collaboration) are conveyed through discourse. Researchers bring extensive ethnographic knowledge to the analysis, often using it to explore how social identities (such as gender or ethnicity) are constructed in interaction. This framework has been especially useful in studying cross-cultural communication and workplace interactions (Gumperz, 1982; Schiffrin, 2003).

Despite these theoretical differences, all of these frameworks share a common tool kit for analyzing discourse, focusing on language use in specific contexts to reveal social meanings and dynamics. Each framework, however, tends to generate different research questions based on its specific theoretical focus.

Developing Research Questions

In the process of developing research questions for a discourse analysis study, the researcher begins with a broad and general inquiry, then refines it based on background reading and the identification of relevant gaps in existing literature. The initial broad question posed was: How do people talk at work? However, upon reviewing existing research, which mainly focused on institutional talk and professional discourse, the researcher sought to narrow the scope and explore everyday workplace communication, particularly in offices and factories.

The researcher then focused on professional white-collar workplaces, an area with which they were familiar. They drew inspiration from previous studies on speech acts like apologies, complaints, and directives (e.g., Clyne, 1994). From this, a more specific research question emerged: How do managers give directives in professional white-collar workplaces?

As the research progressed and a team was formed, additional questions were developed, further specifying the scope of investigation. These included:

How do people make complaints?

How do people refuse to do something at work?

How do people establish rapport with their workmates?

How is power enacted in the workplace?

What is the function of humor in the workplace?

How do people “do leadership” at work?

Are there gender differences in talk at work?

How do people from different ethnic groups talk at work?

Each question reflects a focus on specific discourse features (e.g., directives, complaints, power, leadership, humor) and aexplores how these are performed in different workplace contexts. The evolution of these questions illustrates how a researcher can progressively refine a broad inquiry into focused, context-specific questions suitable for discourse analysis.

Collecting the Data

Discourse analysis can be conducted using various data sources, including written or spoken material, publicly available data (such as newspapers, radio, TV programs, and the internet), corpus data, or data collected by the researcher. The primary advantage of using pre-existing data, such as transcribed spoken material in a corpus, is that it eliminates the need for transcription. However, collecting original data provides a unique advantage, as it allows for a deeper understanding and insight during analysis and interpretation.

To address their research questions, the researchers required authentic workplace talk, which led to the challenge of the observer's paradox—how to capture natural conversation when participants are aware they are being observed. The solution involved recruiting volunteers to record their everyday workplace interactions, which were then transcribed and analyzed. This process, although effective, is complex and involves multiple steps, including obtaining consent, managing data collection logistics, and ensuring the authenticity of the recorded speech.

Some core steps in the data collection process

Make contact with the organization/workplace

The best way is usually through a personal contact.

Identify mutual benefits

Discuss with the most senior person (CEO, Managing Director, boss) what you want to do and identify benefits for the workplace: for example, identifying effective communication strategies.