

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/337858444>

THE ADVANTAGES AND DISADVANTAGES OF CORPUS LINGUISTICS AND CONVERSATION ANALYSIS IN SECOND LANGUAGE STUDIES

Conference Paper · November 2019

CITATIONS

0

READS

2

2 authors:



Cihat Atar

Sakarya University

41 PUBLICATIONS 22 CITATIONS

SEE PROFILE



Cahit Erdem

Afyon Kocatepe University

60 PUBLICATIONS 67 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



NUCASE- Newcastle University Corpus of Academic Spoken English [View project](#)



The Digital Kitchen project [View project](#)

Глухівський національний педагогічний університет
імені Олександра Довженка
Наукове товариство студентів, аспірантів, докторантів і
молодих учених Глухівського НПУ ім. О. Довженка

**ІХ ГЛУХІВСЬКІ НАУКОВІ
ЧИТАННЯ – 2019.
АКТУАЛЬНІ ПИТАННЯ
СУСПІЛЬНИХ ТА
ГУМАНІТАРНИХ НАУК**

*Матеріали науково-практичної
інтернет-конференції
молодих учених і студентів
з міжнародною участю*

25-29 листопада 2019

м. Глухів:
Глухівський НПУ ім. О. Довженка
2019

– Неозначено-особові моделі, у яких особу-діяча не названо, але його можна сприйняти як певне коло людей: *Одержані результати порівнюють із нормативними або аналізують у динаміці.*

– Узагальнено-особові речення, що виражають широкі узагальнення, загальні судження або через форму множини виявляють відтінок співпраці із читачем; активні конструкції з різними формами дієслівних утворень, які передають динаміку думки, точно й влучно відтворюють живе мовлення: *Одержані результати порівнюємо із нормативними або аналізуємо у динаміці.*

Конструкції з предикативними формами на *-но, -то* становлять специфічну рису українського синтаксису і формують ядро пасиву в сучасній українській літературній мові. Опозицію актив / пасив можуть формувати лише перехідні дієслова, пасивними корелятами яких є предикативні пасивні дієприкметники та предикативні форми на *-но, -то*. Традиційно опозицію актив / пасив становлять і дієслівні форми з постфіксом *-ся*. У новітніх дослідженнях конструкції з дієсловами на *-ся* вилучено із парадигми активних-пасивних конструкцій, унаслідок чого вони опинилися на крайній периферії синтаксичної системи української мови.

Поява цих моделей у науковому стилі спричинена впливом на нашу мову російського книжного мовлення. Українська ж літературна мова, сформована на народній основі, здавна відзначалася динамізмом, сконденсованістю думки, і цьому значною мірою сприяла її прив'язаність до активних дієслівних форм. Отож, уживання більшості пасивних моделей є відхиленням від норми літературної української мови.

Література:

1. Куньч З. Уникання пасивних дієслівних конструкцій як елемент культури української мови. *Український смисл: збірник наукових праць / за ред. проф. А. М. Поповського*. 2012. № 1. С. 42–49.

2. Лаврінець О. Пасивні конструкції з дієсловами на *-ся* в сучасній науковій мові: синонімія та паралелізм функціонування. *Українська мова*. 2014. № 1. С. 61–75.

3. Непийвода Н. Ф. Мова науково-технічної літератури: функціонально-стилістичний аспект. Київ, 1997. 303 с.

СЕКЦІЯ 9. ЛІНГВІСТИКА (АНГЛІЙСЬКА, НІМЕЦЬКА МОВИ).

МЕТОДИКА ВИКЛАДАННЯ ІНОЗЕМНОЇ МОВИ

Cihat Atar,

Assist. Prof. Dr., Sakarya University, Turkey

Cahit Erdem,

PhD, AfyonKocatepe University, Turkey

THE ADVANTAGES AND DISADVANTAGES OF CORPUS LINGUISTICS AND CONVERSATION ANALYSIS IN SECOND LANGUAGE STUDIES

Abstract

In this study, Corpus Linguistics (CL) and Conversation Analysis (CA) is discussed as methods of data analysis. Both CL as a quantitative tool and CA as a qualitative tool have their own strengths and weaknesses for applied linguistics studies. So, the tenets of each view are evaluated and the similarities and differences between them are discussed. As for the design of this paper, firstly CL and CA are introduced, and their advantages and disadvantages are mentioned. Then a study which is discussed in the Discussion part is introduced. In the discussion section initially, the similarities and differences between CL and CA are discussed.

Key words: corpus linguistics, conversation analysis, data analysis, second language

From being an approach of a few enthusiasts in 50s and 60s, corpus studies have started to gain more and more supporters recently (O'Keeffe & McCarthy, 2010). Today corpus studies are

used in various areas ranging from dictionary design to formal linguistics. These areas can be grouped under two groups: The first one is for language teaching and learning and the second one is for contributing to academic research in language studies (Hunston, 2002). In accordance with the focus of this assignment, the use of CL only for research in language studies will be discussed.

Applied linguistics benefits from corpus studies a lot. Most of the formal grammar rules in the past were described depending on intuition (Aarts, 1991). This way of studying grammar became popular especially after Chomsky's (1957) claim that grammar rules are abstract rational rules independent of context. However, with the rise of corpus linguistics after 80s, researchers like Sinclair (1991) suggested that empirical data is needed to prove the abstract rules in grammar. As a result, corpora is used to provide objective, quantitative and empirical evidence and it started to serve as an alternative tool for data analysis in formal linguistics.

Advantages and disadvantages of CL in Applied Linguistics studies

Today as a methodological tool CL is an established field in language studies and it provides data to studies and its value is appreciated in the field. However in addition to its benefits, CL also has some limitations.

Firstly CL provides data for language studies by supplying huge amounts of examples about a specific structure and linguists can work out how a structure is used looking at these occurrences. In this way corpus is a really good tool for gathering data about language studies and it provides huge amounts of data which would take a lot of time with traditional data collection methods.

Secondly, in corpus-led language studies, comments and evaluations about language use is not regarded as reliable evidence as they are totally subjective and it is difficult to test them empirically (Kennedy, 1998). In this way, empirical data is at the core of CL and unlike qualitative methods, it abstains from subjective comments or ideas in analysis. The data obtained from corpora is statistically analyzed and concrete evidence is used. Therefore CL depends on objective and empirical data and the data obtained from CL is scientific and it is open to objective verification.

The third advantage of CL is that it provides holistic information about language structures. Thanks to statistical software used in CL, frequencies and statistical information are obtained about certain structures in a study. For example, when a language research about the use of quantifiers is undertaken, CL can supply frequencies of use and researchers can deduce some rules about these structures. In other words, CL does not focus on single contexts, rather it gives an overall picture of a phenomenon thus enabling researchers to have a general understanding about the nature of that specific phenomenon. Hence these results are relevant and reliable in language research because they are drawn out of tens of thousands of usages from real life data. In this regard CL is really successful at large scale analysis.

The final advantage of CL is that corpora include essential information about the data it offers. Most of the modern corpora include information about the socio-economic status, educational background, gender or place of birth of speakers. Moreover, corpora can provide data according to subject, field or genre. For instance, a researcher studying bilinguals can obtain data about bilingual speakers looking at their background or a sociolinguist studying sociolinguistic factors in language can obtain data by looking at the socio-economic status of the speakers in a

corpus. This is another significant advantage of CL because researchers can get relevant data thanks to the information included about the speakers or subjects in a corpus.

As for the disadvantages of data analysis by CL, firstly CL is strictly quantitative and it depends on the description of language via numbers and statistics. Although holistic statistics can be valuable in language research, CL misses some of the most important components of language: the context and structure of the talk among speakers. Language is a social phenomenon and it is difficult to dissociate it from its context (Duranti and Goodwin, 1992). A study done in CL can tell researchers how frequently a structure is used but it cannot tell when, where and how it is used. This is a real challenge for those who want to use CL in qualitative language studies.

In addition, CL disregards intuition as a tool for data gathering on the ground that it is too subjective and it is not testable. In this respect CL has a point but in language research competence as well as performance is significant. Studying the instances taken from natural language performances of a speaker is a reliable method, however researchers such as Nieuwland and Martin (2012) suggest that performance does not equal to competence and studies looking at what is inside speakers' mind is also relevant and indeed it is a must for language studies.

Finally, although CL focuses on large scale analysis and it is good at giving descriptive statistics, it is weak at offering detailed analysis about individual users (Hunston, 2002). What matters the most for CL is what the 'whole' data suggests rather than what individual speakers do. This is of course a problem for CL because it cannot account for individual variation in language use. CL can give information based on the background of speakers and it can make generalizations about certain groups, though. However this is again at group level and it does not look at individual contexts.

Conversation Analysis

Background

CA is an approach to spoken language that considers it as a means of social interaction in contexts of everyday life (Sacks, Schegloff & Jefferson, 1974). It stemmed from the interest in focusing language on its social functions. Initially CA's main focus was on casual conversation but then other situations such as institutional talk were also incorporated into CA. Today, CA is applied in various fields such as sociology, linguistics, anthropology and psychology. CA is especially influential in the fields interactional sociolinguistics, discourse analysis and discursive psychology (Heritage & Clayman, 2010).

The underlying philosophy of CA is that speech in social contexts is not static. The social contexts in which communication occurs are constantly shaped by how participants use language and by their management of turn-taking, sequencing of acts or conversation openings and closings (ibid). In addition, CA studies language at micro level and rather than having an initial hypothesis, the aim is to let the data speak for itself.

Advantages and disadvantages of CA in Applied Linguistics studies

The first advantage of CA is its sensitivity to meaning and context. How turns are managed, why language users' language use (such as intonation, lexical choice or syntax) is adjusted to the current communicative situation or the sequencing of events all mean something in conversation (Fetzer, 2004). This aspect of language cannot be understood via quantitative analysis and a qualitative method such as CA is needed to have a closer look at this aspect of language.

Secondly, CA offers detailed descriptions from data. This enables researchers to have a deeper understanding about the language used in a context. Unlike quantitative methods, CA investigates data more closely and qualitatively rather than giving some numbers and statistics about the data. In this respect CA provides a richer and deeper understanding of research data.

Another property of CA is that it looks at data without having any prior aim. The reason for this approach is having an open mind while doing a research. When a researcher has some prior decision in his/her mind about his/her study, it is possible that the way s/he approaches the data will be affected. Therefore CA suggests that researchers should let the data speak for itself (Lerner, 2004). This allows researchers to analyze data more objectively and they can recognize very unique properties in it.

Finally, the focus of CA is the naturally occurring language in comparison to Chomskyan approaches that suggest studying the idealized competence of speakers (Atkinson and Heritage, 1984). Chomskyan approaches in applied linguistics field suggest that performance of speakers is a bad image of what is inside their mind. The adjustments in language in accordance with different contexts are seen as deviations from the norm. Therefore they suggest that the competence of speakers should be tested context-free. However, CA assigns an important role on natural speech because language is seen as a social phenomenon that is affected by context, prior knowledge of speakers and power relations. Namely, CA is directed towards a more naturalistic account of language use.

As for the disadvantages of CA, firstly and most importantly, the findings of CA cannot be generalized to other contexts (Walsh *et al.*, 2011). The findings of a study in a certain context are valid only in that specific context and they cannot be used to make assumptions about other contexts. This is a problem for CA because most studies in science aim at not only obtaining findings but also being able to extend those findings to other problem areas.

Another disadvantage of CA is that it can handle little data in a given study. Owing to its nature, CA studies focus on very little data. The problem with being able to analyze little data is that it diminishes the reliability of studies. In other words, when only a little bit of data is analyzed in a study, there will be huge amounts of other data which may also be studied. In this respect, the nature of CA practice prevents it from studying large enough data to obtain reliable and generalizable data.

Discussion

As discussed above CL and CA have both strengths and weaknesses. Firstly, CL deals with large amounts of data while CA can only deal with a small amount of data. Being able to analyze more data enables CL to make generalizations about study results. Moreover, studying enough data increases the validity and reliability of a study. However, despite CA's inadequacy in terms of the data it can analyze, CA also has some merits. CA studies only small amounts of data but it offers a deeper understanding into the data. It also focuses on meaning and context unlike most CL studies and this enables CA to evaluate the data much more accurately. In this respect, while CL is able to analyze a big amount of data and offer more reliable and generalizable study results, CA provides a deeper understanding into the data in a study.

With respect to the generalizability issue in the previous paragraph, CL seems to be more advantageous. The reason is that the aim of the study asks for a generalization about a certain group, the Turkish bilinguals. Besides in order to have a reliable study, enough amounts of data should be analyzed. Thus, CL can both give the general picture about a phenomenon and it can

also demonstrate it with many examples from a corpus. On the other hand, for the topics that are really susceptible to the context in which it is used, a qualitative account may be better. In this respect CA can give a closer account of the data and it can analyze the contexts much better. This will be of great use because CL may possibly include many other different details in the data which are not directly related to the topic in focus. In order to clearly distinguish different phenomena, CA can offer a micro analysis which includes only the relevant examples. Consequently, the validity of second language studies can be increased via the incorporation of CA into CL.

Secondly, CL can be used as a quantitative tool with statistics and the empirical data it provides. However CA produces qualitative results and it mostly relies on the comments and evaluations of the researcher. Generally when the term science and research is mentioned, quantitative methods seem to be more appropriate. The reason is that their studies depend on empirical data and processes like proposing a hypothesis and testing the hypothesis against the data sound more objective and scientific. However social sciences may work in a different way compared to the fields like mathematics or physics. Therefore, the fact that CL offers empirical and measurable data does not mean that it is better in language studies. Rather the important thing is to offer an appropriate analysis for a certain study. So, CL is more advantageous in the research that aims at offering empirical data and more importantly it can test the hypothesis of the study.

Next, a similarity between CL and CA is that not only CA but also CL puts an emphasis on the use of natural and daily speech in language studies. This is in contrast to rationalist approaches in language studies which suggest studying the competence of speakers. As discussed in the Literature Review part, studying the performance of people is a much more realistic and valid way in Applied Linguistics as it can be easily tested and recorded. However depending on context, sometimes testing competence of people may also be important. But both CL and CA cannot provide data about the competence of people. CL would be of great use when many examples from real life are needed as it can show how the phenomenon is used. This is better than rationalist tests like grammaticality judgement tasks (GJT) because CL shows direct evidence from how people use it but GJTs can only show what people 'think' about the grammaticality of a structure. In addition, CA can also contribute by providing highly context-specific instances and in fact, it is the context which makes most of the interaction meaningful. Consequently, CL and CA together favor natural data in second language studies and this is appropriate if a naturalistic approach is sought.

Finally, one peculiar characteristic of CA compared to quantitative methods is that it looks at the data without having any prior aim. The idea of letting the data speak for itself may enable researchers to look beyond their hypothesis. When researchers have a hypothesis in their mind, they may be inclined to see exaggerated links with their hypothesis and the data which results in unreliable findings. In addition, having a prior decision may prevent researchers from seeing some other very precious aspects of language which is not the focus of the study and researchers lose the chance of unearthing those aspects. Quantitative methods such as CL on the other hand first have a hypothesis and then try to verify or refute it. In this respect, CL is more appropriate when there is a need for empirical data of many instances. Consequently, this type of a context is challenging for CA.

Conclusion

As seen in both the Literature Review and Discussion part, both CL and CA have their own advantages and disadvantages. While CL can offer a broad picture of language, with empirical data and many examples from corpora, CA is good at providing rich details from a data in a qualitative manner and it takes the importance of context and meaning into consideration thus making language studies better. When considering the advantages and disadvantages of CL and CA, it can be concluded that CL is more appropriate in quantitative studies. CL can be used to provide the description of a phenomenon and this can be verified via empirical examples. However, as discussed in the Discussion part, CA can also provide essential information and it can really help increase the reliability and validity of a study through the detailed analysis of context and meaning. Therefore it can be concluded that CL and CA should be used in a complementary relation rather than seeing them as two competitive and exclusive tools of data analysis.

References

- Aarts, J. (1991). Intuition-Based and Observation-Based Grammars. In Aijmer, K and B. Altenberg (Eds.), *English Corpus Linguistics*. London: Longman.
- Atkinson, J. Maxwell and Heritage, J. (eds) (1984). *Structures of Social Action: Studies in Conversation Analysis*, Cambridge: Cambridge University Press.
- Chomsky, N. (1957). *Syntactic structures*. Mouton: The Hague.
- Cook, V. J. (2002) Background to the L2 user. In V.J. Cook (ed.) *Portraits of the L2 user*. Clevedon: Multilingual Matters. (pp. 1-28).
- Duranti, A., & Goodwin, C. (Eds.). (1992). *Rethinking context. Language as an interactive phenomenon*. Cambridge: Cambridge University Press.
- Fetzer, A. (2004). *Recontextualizing context*. Amsterdam: John Benjamins.
- Heritage, John and Steven E. Clayman (2010). *Talk in Action: Interactions, Identities and Institutions*. Boston: Wiley-Blackwell.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Kennedy, G. (1998) *An Introduction to Corpus Linguistics*. London: Pearson Education Limited.
- Kerslake, C. and Goksel, A. (2005) *Turkish: A Comprehensive Grammar*. London: Routledge.
- Lerner, G. H. (ed.) (2004) *Conversation Analysis: studies from the first generation*. Philadelphia: John Benjamins Publishing.
- Lewis, G. (2000) *Turkish Grammar*. Oxford: Oxford University Press.
- Nieuwland, M. S. and Martin, A. E. (2012). If the real-world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition*, 122, 102-109.
- O'Keeffe, A. and McCarthy, M. (2010) *The Routledge Handbook of Corpus Linguistics* Routledge, Hoboken, ISBN: 9780203856949
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696-735.
- Sinclair, J. M. (1991) *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.

Walsh, S., Morton, T. and O'Keeffe, A. (2011) Analysing university spoken interaction: A CL/CA approach. *International Journal of Corpus Linguistics, Applying Corpus Linguistics*, pp. 325-345(21)

УДК 81 – 11

Ломакіна Д.С. -
студентка 21 ДО групи
ППК Глухівського НПУ ім. О. Довженка
Науковий керівник:
Дворянова Т.О. - викладач
Професійно-педагогічного коледжу
Глухівського НПУ ім. О. Довженка

ЗАПОЗИЧЕННЯ В АНГЛІЙСЬКІЙ МОВІ ТА ЇХ ПЕРЕКЛАД

У складі будь-якої мови присутні не тільки споконвічно рідні слова, а й численні запозичення з інших мов. В одній мові їх більше, в іншій - менше, але вони є завжди. Запозичення - це позитивний процес для мови. Таким чином збагачується словниковий запас, з'являються нові елементи для словотвору, вводиться лексика, що належить точній термінології. Кількість запозичень в англійській мові досить велика в порівнянні з іншими мовами.

Підраховали, що на запозичення в англійській мові припадає близько 70 відсотків лексики, а решта, що залишилася і становлять споконвічно англійські слова. Причина цього криється в історичному розвитку англійської мови. Численні іноземні завоювання Британських островів, розвиток торгівлі, вплив культури континентальних країн - все це сприяло появі і закріпленню запозичень в англійській мові [2, с. 56].

Так англійська мова прийняла багато слів з японської, італійської, німецької, французької, китайської, іспанської та навіть перської мов. Але, давайте почнемо з самого початку.

Неможливо обговорювати англійську мову, не згадуючи вплив Давньонорвезької мови - масштабне впровадження скандинавських слів у мову, що зараз відома як стара англійська (Old English). Під час нормандської окупації, більш ніж 10000 нових слів були введені і більшість з них знаходяться в експлуатації [4, с. 25].

Це такі слова: cake (торт), fog (туман), die (померти), happy (щасливий), husband (чоловік), leg (нога), mistake (помилка), window (вікно).

У той же час, багато старих французьких слів стали вводитися в англійську оскільки французька була встановлена в якості мови суду. Вплив французької легко впізнається сьогодні. 30% лексики походить від французької мови - liberty, justice на противагу freedom і fairness, а також деякі прості слова: machine (машина), garage (гараж), terrain (рельєф), collage (колаж) [3, с. 86].

Англійська використовує багато слів, запозичених з Японії. У той час як більшість з них позначають щось конкретне з японської культури і способу життя (наприклад, бойові мистецтва), інші з'явилися через популярні бренди.