Linguistic reunification of Germany: an eclectic evaluation after 30 years

Esther Looijen
s1010411
08-02-2022
Linguistics
Hans van Halteren

Preface

Before you lies the bachelor thesis "Linguistic reunification of Germany: an eclectic evaluation after 30 years". It is the result of my love for the technical side of linguistics, discovered during the bachelor Linguistics at the Radboud University Nijmegen for which this dissertation is written, and an interest in the technical workings of the German language as it is my mother's native tongue.

Together with my supervisors, dr. Hans van Halteren from the Department of Language and Communication, and drs. René Gerritsen from the Modern Languages and Cultures department with a specialisation in German Language and Culture, the research question was formed and the methods were decided upon. Contact with the *Institut für Deutsche Sprache* in Mannheim was initiated by dr. van Halteren, who had worked with the institute before and was pleased to have an opportunity to rekindle contact; the planning of subsequent Zoom-calls and the writing of follow-up emails gave me an excellent chance to learn to work with large institutes.

While I would have liked to work with the IdS in person, unfortunately Covid-19 threw a spanner in the works, as it has for many theses and other things in life. It dramatically prolonged the writing process: many people willing to help acquire the dataset were ill at one point or another. Myself and my supervisor were also out of the running for a while, and with everyone at home, some peace and quiet was hard to come by (not so much in my house, but more the five kids next door). Nevertheless, in the end, hard work prevails; although it was later than expected, this thesis is finished after all.

I would like to sincerely thank both of my supervisors for not only answering my every question, but also providing support, insights, and information above and beyond what was required. Additionally, I would like to thank my parents for remaining patient with me; many lunches and dinners were filled with me thinking out loud about this research and asking their opinions. Lastly, my fellow student and friend Ellen deserves a particular note of thanks: you were always there when I needed a listening ear or some wise words.

I hope you enjoy reading.

Esther Looijen

Rheden, February 8, 2022.

Contents

Chapter 1: Introduction

It has been 32 years since the Reunification of West and East Germany removed a major political, social, cultural, and partially physical barrier between two parts of Germany that had grown apart during the separation. In some areas, this was very apparent: both parts immediately and clearly aligned themselves with different political ideologies. Other disparities were harder to see, especially from the outside. One of those was the fact that East and West Germany gradually grew apart linguistically. The social, cultural, and political differences manifested themselves in the official, political language as well as in everyday language; in East Germany, especially the latter was hard to see or research, because the political system actively tried to repress it in favour of a unified East German language. Nonetheless, it has been painstakingly documented and researched by scientists as soon as they became aware of the ever increasing differences in the 1960s; this continued well into the 1990s, after the Reunification. Since then, however, relatively few investigations into the current state of differences between East and West German language have been conducted. Research done shortly after the Reunification suggested that the language in the two parts was growing closer again, much like what happened in other areas; owing to the lack of recent research, however, it is not known definitively if this process was completed.

As soon as scientists became aware of the differing tendencies of language in both parts of Germany, a closer look was taken at how they came to be, what factors influenced the diverging, and how the differences manifested themselves. A concise overview will be given here; the complete literature review can be found in Chapter 2. As mentioned, the political differences between East and West were a very public and obvious factor that played into the linguistic differences. The socialist politicians in the East saw language as part of their identity (Bock et al., 1973); as a result of this, an "official" version of East German was spoken by the government and forced upon the East German population. An unofficial "everyday" version of the language emerged as well; however, researchers often did not have access to this - it was spoken mainly in private settings, people were careful to hide it as it was forbidden by the government. Additionally, perhaps subconsciously, researchers were affected by the politics in the part of Germany they were from. Governments tended to blame each other for "ruining the German language" (Kreutz, 1997), and this world view, possibly combined with pressure from the government to publish favourable results, might have blinded (mainly West German) researchers to the presence of something other than the official East German language (Hellmann, 1980). After the Reunification, the existence of the "everyday" East German came to light and it was well researched, for instance by Schlosser (1991) and Kreutz (1997).
Lexical differences were the most common linguistic differences between East and West German. For instance, again under political influence, some Russian loan words started to embed themselves into East German (Hellmann, 1984), and some English words into West German (Uchimura, 1983). Around the same time as these politically influenced words were identified, a classification system was made for the lexical differences between East and West German language (Hellmann, 1980). Examples of categories were words specific to only one part of Germany, words that have different meanings in both parts, different words for the same thing, or words that have a more positive rating in one of the two parts. In this same study, Hellmann also noted that no area of life seemed to be exempt from lexical differences. This was attested by Pankanin (2017): words defined in the German Duden-dictionary as being

exclusively or mostly GDR-used were found in both official areas like politics and work, but also in education, sports and everyday life. This was, thus, still the case many years after the Reunification.
Other linguistic areas were barely or not at all affected by the division of Germany; this was established, for instance, for syntax and morphology (Bock, 1977; Folsom & Rencher, 1977; Hellmann, 1980). Every now and then, some small and very specific syntactic or morphological differences were found (Hellmann, 2008); however, it was generally accepted that deeper layers of the language like grammar or syntax change more slowly than the lexicon (Steffens, 2009) - the latter is directly influenced by the world around the speakers.
Persisting even after the Reunification were pragmatic differences and communicative issues. This was partially caused by the insecurity of East Germans who had had to hide their everyday language for so long, and thus were unsure how to talk to "outsiders" (Markkanen & Schröder, 1996). They didn't want to be too disrespectful and therefore adjusted their language to sound less harsh. On the contrary, they were disproportionately harsh towards themselves (Schlosser, 1991). This continued for at least two decades after the fall of the Berlin wall and the Reunification: several researchers, among them Plewnia & Rothe (2009), still identified a *Sprachmauer* ("speech wall") in both West and East Germans' communicative attitudes.

The question now remains: have the linguistic differences between East and West Germany diminished or disappeared today? After all, most of the differences were lexical; these tend to change relatively rapidly and are heavily influenced by the world around the language users - it stands to reason that these lexical differences would mostly have disappeared since the two halves of Germany were reunited. Conversely, at least pragmatic and communicative differences have been found well after the Reunification; if they existed in 2009, some differences in language between East and West Germany might still be present today. This study will therefore aim to answer the question:

To what extent are linguistic differences present in East and West German newspapers?

The expectation is that there are few, if any linguistic differences today. Especially lexical differences are not expected to be present to the degree they were before 1990 - seeing as lexical differences are quickly and heavily influenced by the world around the language users, and that Germany has been unified for over 30 years, the largest factors influencing the previously present lexical differences should have disappeared, along with the lexical differences themselves. Other linguistic differences like syntactic or morphological differences were barely or not at all present even when the two parts were separated - these are not expected to be present either.

In order to find an answer to the research question, three different methods will be attempted. All methods will be based on a dataset consisting of all 2019 editions from nineteen East and West German newspapers, which is taken from DeReKo, a German corpus consisting of newspapers, magazines, internet pages, and many more sources. The first method, found in Chapter 3, will be based on literature describing past linguistic differences. A list of words that definitively differed in usage between West and East Germany at one point in time between 1945 and 1990, is compiled; in order to compare the past differences to today's, the same words will be looked at for the current dataset. The second method, described in Chapter 4, is based on research done by Hellmann (1984) - a statistical method was used to determine which words were used significantly more in one of the two parts. This research is replicated in the current study to see which words differ significantly between the two parts today. The last method,

found in Chapter 5, uses a different, data-driven statistical approach - a newer method that was not used in earlier research - in order to see what differences a more modern approach unearths. All these methods are accompanied by an extensive literature review in Chapter 2, and are concluded with study limitations and a final conclusion in Chapter 6.

Chapter 2: Literature review

In this section, a brief overview of several aspects of the linguistic differences between East and West Germany will be given. First, a few cultural-political factors that had a direct influence on the German language in either part of Germany are discussed. This is necessary to gain an understanding of the origin of these differences. Section 2.2 talks about lexical differences between East and West German; these are divided into several broad categories and a few specific areas. Sections 2.3 and 2.4 encompass non-lexical differences; syntactic, semantic, stylistic, and pragmatic differences are discussed.

2.1 - Cultural-political background

In the early 1960s, the linguistic differences between East and West Germany became apparent to linguists from the GRD. Sporadic papers on the matter began to appear, but they mostly lacked sufficient material for a solid research basis and showed political-polemical intentions (Hellmann, 1980): the differences in language were seen as a problem for which the finger was pointed at perceived shortcomings or wrongdoings on the other side. This was also attested by Kreutz (1997): both sides accused each other of ruining the German language with either their imperialist or socialist tendencies. These tendencies impacted the respective lexicons and the living environment, which in turn had an effect on the language people use. In the East, the socialist government aimed to align themselves with like-minded countries, the largest one being Russia - the Eastern lexicon therefore began to contain Russian loan words like *Brigade* (ru. *brigada*, "brigade") (Hellmann, 1984). The Western government's imperialism was more in line with American politics; as a result, many English words were borrowed (Uchimura, 1983; Jarausch, 2012). The borrowing of English words was not exclusive to West Germany; some English words did find themselves borrowed by German speakers in East Germany as well.

The existence of political influences also meant that research, especially before the Reunification in 1989, was not independent - scientists were, maybe subconsciously, influenced by the political ideology that was present in their part of Germany (Hellmann, 1980). Later on, closer to the reunification, this finger pointing was slowly replaced by a certain level of tolerance for "the other side" and by a desire to work together. Research became less biased and more empirical; a better research environment was created. At congresses held after the Reunification where both East and West German researchers were present, active interference from the East German government was denied by the East German scientists (Schlosser, 1991), but the subtle, perhaps subconscious manipulations by the researchers' current world views as found by Hellmann (1980) may well have played a part nonetheless.

The political situation not only influenced research, it also affected the German language itself. Hellmann (1980) identifies a political-ideological aspect as well as national, social, and communicative aspects to the communicative problem that arose between East and West Germany. In the East, marxism-leninism caused a class conflict whereas in the West, a general anticommunist attitude existed (Bock et al., 1973); this in turn caused a rift between them about the "real" meaning of words (see section 2.2). Besides these ideological differences, there were dissimilarities in the sense of nationalism: whereas the West was inclined towards a singular Germany with a single language, the East saw themselves as an up-and-coming "mostly German" socialist state where their language was part of their identity. Hellmann further noted a communicative problem that was not well attested at the time of his writing: some researchers say that communication between people from the East and West was largely effortless, whereas

others say there was a certain difficulty that differed per subject or situation. This will be further touched upon in section 2.4. Lastly, Hellmann writes of a social aspect: a discrepancy between everyday and political language in the East. The latter was much more interesting to researchers from the East. This might have, however, been because of lack of knowledge about day-to-day life in the GDR - the socialist government actively blocked access to "the common folk" and their language use.

Under the socialist regime in East Germany and the rules it enforced upon its citizens, a language split thus emerged: on the one hand, there was the "official" German that was used by politicians and which was forced upon the population by the government. This was also the German used in newspapers and official documents and correspondence. On the other hand, there existed a form of German used by regular people in their everyday lives. The existence of this double-speak was quite well-known; it has been discussed to varying degrees, for instance in Hellmann (1980), Schlosser (1991), Von Polenz (1993), Kreutz (1997), and Jarausch (2012). However, as can be seen from the years of publication from most of these papers, it only came to light after the Reunification when researchers gained access to the general public.

The East German political language was characterised by more nominally inclined sentences, chains of genitival or prepositional attributes, and excessive use of adjectives and fixed expressions (Hellmann, 1980; Jarausch, 2012). Examples of the latter include adjective-noun pairs like *sozialistische Menschengemeinschaft* ("socialist human community") or verb combinations like *einholen und überholen* ("to catch up and overtake"). This resulted in hollow phrases like *die ökonomische Hauptaufgabe in ihrer Einheit von Wirtschafts- und Sozialpolitik* ("the economical main task in its unity of economy and social politics"). The general public distanced themselves from monstrosities like this and called them *Parteichinesisch* ("political party gibberish"), *Kaderwelsch* ("party official gibberish", a portmanteau of *(Partei)kader* - "political party officials", and *Kauderwelsch* - "gibberish"), or *Hoch-DDRsch* ("High-DDR-ish", a play on *Hochdeutsch* - "High German"; German without accent as spoken by educated people, but sometimes seen as pretentious). The negative stance towards this "official" East German language, however, was not well known outside East Germany until after the Reunification because of the government's suppression. After 1989, the "real" language of the East German public came to light, although mostly in the form of anecdotal evidence: it was full of puns, ironies, and parodistic words (Von Polenz, 1993).

Schmidt (2009) gives a few examples of how mainly East Germans would play around - in private, of course - with the formulaic language they were exposed to by their government. After 1990, this wordplay came to light when it was published in newspapers; the first sentence of the *Communist Manifesto* by Karl Marx and Friedrich Engels, for instance, was something every East German learned in school. *Ein Gespenst geht um in Europa - das Gespenst des Kommunismus* ("A spectre is haunting Europe - the spectre of Communism"). The last word was replaced with words like *Optimismus* ("optimism") or *Gleichgültigkeit* ("indifference"), or the whole sentence was adapted to become, for instance, *Ein Gespenst geht um im geeinten Deutschland: "Ost-Identität"* ("A spectre is haunting unified Germany: "East-Identity").

The obscured nature and delayed discovery of the existence of this everyday language meant that it is difficult to make a historical overview of East German as spoken by the general public; there is little to no written material available from before 1989, as public texts were often monitored by the government and therefore necessarily in "official" East German.

The final aspect of the cultural-political differences that should be mentioned is a discussion between East and West German researchers: the question if East and West German were two different languages, or if they were two dialects or variants of a single language. The answer to

this question helps to put the "East-West language issue" (*das sprachliche Ost-West-Problem*, as Dieckmann coined it in 1967) into perspective. Because the East German government wanted to establish their own German socialist state, they saw having their own language as part of their goals (Uchimura, 1983). This was reflected in their policies, so East German researchers may have been influenced by this - maybe subconsciously. Especially when the changes within the German language were first noted in the early 1960s, however, many (predominantly West German) scientists were of the opinion that German was still one singular language. They, too, were driven by ideological motives - they criticised the communist ideology from the East, and therefore their drive to create a new German socialist state with its own language (Kreutz, 1997). Furthermore, the only known differences between East and West German were lexical in nature, whereas all other linguistic aspects seemed unaffected. Even the East German researcher Victor Klemperer, who meticulously documented the smallest details about the German language under Nazi reign (he called it *LTI - Lingua Tertii Imperii*) and came to the realisation that marxism-leninism was influencing that German in such a way that another form of German (*LQI - Lingua Quatri Imperii*) emerged, wrote in his diaries that he saw little to no differences between LTI and LQI (Young, 2005). As time passed and more research was done, the discussion shifted from a question of two languages to a question of degree of communicative problems; East and West German were increasingly seen as mere variants of a single language much like Austrian was a variant of German (Hellmann, 2008; Von Polenz, 2009), but there was a certain difficulty in communication between the two parts of Germany. The communicative problems will be further touched upon in section 2.4.

2.2 - Lexical phenomena

As mentioned in the previous section, the existence of lexical differences between East and West German was discovered relatively early on: Moser (1954) ascertained that the Eastern and Western lexicons were drifting apart somewhat. He downplayed the degree and impact of the divergence, however; the lexical differences were reportedly very small, and no other linguistic areas were affected. This stance was upheld until the 1970s (Kreutz, 1997), when linguists realised that the lexical differences alone were significant enough to warrant in-depth research. Various corpus studies (e.g. Hellmann, 1984) performed statistical analyses on the differences between Western and Eastern newspapers, and found significant and systematic differences between words used in both parts of Germany. Until this time, Western researchers often thought it was a degradation of the German language in East Germany that caused the differences and vice versa. It turned out, however, that the lexical divergence found its origin in both East and West: a screening for new words in 1960, as described in Dieckmann (1967), yielded numerous neologisms, but the new words found in West Germany differed from the ones found in East Germany.

A common classification system for different types of lexical differences was proposed by Hellmann (1980), and later expanded upon in Hellmann (1984). It differentiates between *Lexemspezifika* ("lexem specifics"), *Bedeutungsspezifika* ("meaning specifics"), *Bezeichnungsspezifika* ("designation specifics"), *Gebrauchsspezifika* ("usage specifics"), *Häufigkeitsspezifika* ("prevalence specifics"), and *Wertungsspezifika* ("evaluation specifics"). *Lexemspezifika* are the most numerous type, and encompassed those words that were only attested in one of the two parts of Germany - they were either merely used in citations or discontinued on the other side.

Bedeutungsspezifika are words that have the same phonological properties, but a different meaning on either side. This happened for instance with the abbreviation *APO*, which meant

Abteilungs-Partei-Organisation in the East and *Außerparlamentarische Opposition* in the West. Another possibility is the addition, loss, or substitution of one or more meanings of a word on either side. The existing word *Brigade* ("brigade"), a military term for the smallest unit of soldiers, gained a new meaning in East Germany under the influence of Russian - a socialist brigade, the smallest unit of workers for a given task in a production process.
Bezeichnungsspezifika are words with (slightly) different phonological properties, but the same meaning. Examples of this, with the Western lexem on the left and the Eastern on the right, include *Plastik / Plaste* ("plastic"), *Staatsangehörigkeit / Staatsbürgerschaft* ("citizenship"), or *Arbeitnehmer / Werktätiger* ("employee"). Like the *Bedeutungsspezifika*, these differences could be caused by retention of an older meaning: whereas the West German word for the lowest rank for a general was *Generalmajor*, the same meaning in East Germany was attributed to *Brigadegeneral*, again under Russian influence.
Unlike the previous three categories, the next three are more based in sociolinguistics; researchers who based their work on this classification sometimes omit one or more of these categories for the difficulty in measuring them, or for the fuzziness of the boundaries.
Häufigkeitsspezifika have different frequencies of appearance in East and West Germany. They are attested in both parts, but are more often used in one of them. For East Germany, words like *sozialistisch* ("socialist", adjective), *Produktion* ("production"), or *wir* and *unser* ("we" and "our") appear more often than they do in the West; on the other hand, words like *freiheitlich* ("free", adjective)*, Partnerschaft* ("partnership")*,* or *dynamisch* ("dynamic", adjective) are more frequent in the West than in the East.
Wertungsspezifika have different positivity ratings in both parts of Germany. This is, however, independent of the meaning of a word in East or West Germany - while *demokratisch* ("democratic") has different definitions, it receives the same positivity rating. The same cannot be said for words like *Kommunist* ("communist")*, Klassenkampf* ("class struggles")*, christlich* ("christian") or *idealistisch* ("idealistic"); the first two receive a positive rating in East Germany and a negative one in West Germany, while the opposite is true for the latter two.
Gebrauchsspezifika refer to the differences in usage of available words and style possibilities. This is closely tied to societal norms - they dictate how one should behave in society, and therefore influence how people use language. This is harder to measure, and examples are not easily written down in one or two words; nonetheless, the existence of these differences should be noted.
As Hellmann (1980) further noted, no subject or life area was exempt from these linguistic differences. Lerchner (1992) determined that the combination of different political, economical, and social circumstances in East and West Germany caused different "communication cultures"; therefore, it stands to reason that all areas of life where communication is necessary, i.e. every single area, would be affected. However, some areas were affected more severely than others, according to Hellmann; this is also logical, seeing as the three areas in which Lerchner found major differences would be most prominently affected. Pankanin (2017) analysed 328 words marked in the German Duden-dictionary as either exclusively GDR-used, mostly GDR-used, or formerly used in the GDR. Most of these are now archaisms, but it still paints a compelling picture in regards to how widespread the use of these words was in areas of life in East Germany. The results are summarised in the illustration below (Pankanin, 2017: 118); *image omitted due to copyright issues*.

Diagramm 2. Der Anteil der semantischen Gruppen am gesamten Untersuchungsmaterial.

Besides politics and law, to which 15% of the analysed words belonged, and other more official areas like economics (13%) and work (17%), a fairly large share of GDR-specific words was to be found in everyday life (12%), education (12%), and communal activities like sports (16%). This shows that the linguistic differences between East and West Germany weren't restricted to the "official" East German language, but were spread throughout the entire society.

The question now, of course, remains to what extent these differences might still be present today. According to Steffens (2009), shortly before and after the Reunification is when typical East German words or specific East German meanings of words started to disappear; they were replaced with West German words or meanings. Hellmann (1990), however, compiled a small glossary of words used in everyday language (i.e. non-political words) that he considers to be *Wende-resistent* (literally "Turn-resistant") - they survived the turbulent time of the Reunification. These words are later revisited in Hellmann (1997) and Steffens (2009); in 1997, they were still used frequently, perhaps under the influence of nostalgia or a growing sense of self-confidence among East Germans. In 2009, however, it seemed that many words were disused, particularly in public, official contexts. Privately, many were still used occasionally, but with a much lower frequency than before. Steffens attributes this to a difference in prestige between East and West German; West German had higher prestige, so over time, East Germans adapted their language to resemble the higher prestige variety.

Steffens (2009) identifies another set of East German words that are still widely used; they are mostly *Bezeichnungsspezifika* that don't stand out because of their outlandish form. For instance, East German *alleinstehende Mutter* (literally "alone-standing mother") is widely used next to West German *alleinerziehende Mutter* (literally "alone-raising mother") as a pair of words for "a single mother". These two variants differ relatively little from one another and are therefore both still used, according to Steffens; on the other hand, East German *Jahresendflügelfigur* (literally "end-of-year-wing-figure") as a word for "Christmas angel", a type of decoration for the Christmas tree, lost to West German *Weihnachtsengel* (literally "Christmas angel") because of its weird form.

The presence of East German words in texts is analysed as follows by Schmidt (2009): there are plenty of words that, while much older than the GDR, are associated with East Germany. If several or many of these words are present in a text, one can safely assume that the text originates in East Germany; it is not a single word that makes a text East German, but a set of specific meanings (*Bedeutungsspezifika*) of words that occur together. Overall, newspapers from East and West Germany only showed a difference in vocabulary of merely three to four percent, but this was enough to make texts differ in the way they looked at the world, according to Jarausch (2012). Little research regarding this was done in the last ten years (between 2009 and 2020); therefore, it is hard to say if the three to four percent difference is still an accurate number, and if the differences are still notable enough to be able to conclusively attribute a text to East or West Germany.

2.3 - Morphological, syntactical, and semantic phenomena

Unlike the abundance of lexical differences, quite few morphological, syntactical, and semantic differences have been attested over the years. As Bock et al. (1977) and Hellmann (1980) summarised, syntax and morphology show the same patterns and tendencies in East and West Germany. Most researchers are of the same opinion; whereas lexical changes are usually directly influenced by the world around language users and therefore have the potential to happen rapidly, other areas like grammar, syntax, or the phonetic inventory tend to change at a much slower pace (Steffens, 2009). These relatively stable aspects have therefore often been left out in large studies of the German language - East and West Germany were divided for a rather short period compared to their extensive history, so if there were any linguistic differences at all, they were assumed to be in the lexicon. Nonetheless, some papers have looked at the possibility of syntactic differences; Folsom & Rencher (1977), for instance, found no significant syntactic differences in East and West German prose texts.

Every now and then, studies would find some very specific differences in the morphologic or syntactic area. It was not enough to deter from the general consensus that these areas behaved in largely the same way in East and West German, but these differences should nonetheless be mentioned at this point. Hellmann (1980; 2008) found some stylistic differences that are technically syntactic in nature. For example, he found a much higher frequency of the masculine and neuter genitive article *des* - East German official language was full of strung together genitive constructions like *der Stellvertreter des Vorsitzenden des Ministerrates der Deutschen Demokratischen Republik* ("the representative of the chairman of the Council of Ministers of the German Democratic Republic"). A collectivistic "we" style was also present in East Germany - words like *wir, us,* and *unser* appear much more frequently in the East German newspaper than they do in the West German one that Hellmann examined - he followed this with the observation that the opposite was not true, i.e. the West German newspaper did not necessarily contain more individualistic "I"s than the East German newspaper.

Besides those differences, it was also noted that East German used adjectives and nouns more frequently. Those adjectives seem to entail mostly ideologic words in conjunction with a noun - these together usually formed fixed expressions and set phrases, by which the East German language was plagued. Reich gives examples like *demokratischer Aufbau* ("democratic buildup") and *marxistische Ethik* ("marxistic ethics"). Furthermore, Hellmann (1990) found through statistical analysis of newspapers that East German had a predilection for superlative adjectives - adjectives that signalled a high degree of something. Examples include *allseitig* ("all-round"), *konkret* ("concrete"), *breit* ("broad"), and *komplex* ("complex")." West German newspapers, on the other hand, showed a much higher frequency of more moderate adjectives like *einerseits* and

andererseits (“on one hand”, “on the other hand”), *etwa* (“approximately”), *vielleicht* (“maybe”), and *sicherlich* (“surely”). Hellmann (1990) already saw this kind of adjective appear more in East German newspapers very shortly after the start of the Reunification.
Another type of stylistic difference, a few very specific cases of differences in word formation, is mentioned by Hellmann (1980); in West Germany, a substantial amount of compound words starting with *Euro-* (for “Europe”) were attested, whereas in East Germany, the same was true for several compounds starting with *Inter-*.
In general, East German showed the most divergence from what was considered “standard German” - the stylistic differences described above, for example, are mostly differences that only occurred in East Germany (Hellmann, 1980); however, the divergence in West German language should not be neglected. Hellmann gives an example of an exclusively West German pattern of word formation: *sit-in*. This shows the influence of English in West Germany - it is an English word that took the place of German *Sitzstreik*, and sprouted many new related words like *teach-in* - adapted from English, but only accepted in the West.
A semantic problem arose from the existence of *Bedeutungsspezifika*. As Von Polenz (1993) found, these words, that had the same orthography but a different meaning in East and West Germany, could be entirely unproblematic on one side; on the other side, however, they could have an entirely different meaning, connotation, or implication. Words like *Warteschleife* (literally “waiting loop”, “queue”) or *Rechtsstaat* (“constitutional state”) were nothing special for West Germans, but East Germans would use them exclusively in an ironic context. Von Polenz saw these as a type of ‘false friends’; these would, in his and other researcher’s eyes, be significant hurdles on the road to reunifying Germany’s language.

2.4 - Pragmatic phenomena and communicative issues

When the dust had settled after the destruction of the Berlin Wall and the Reunification, there was an awkward atmosphere between East and West Germans. It became apparent that they had issues communicating with one another; they used different words, had different communicative styles, and especially the East Germans, after having to hide their everyday language for so long, had trouble figuring out how to conduct themselves in conversations with “outsiders”.
Antos & Schubert (1997) discuss what part of the communicative issues between East and West Germany can be attributed to linguistic problems - they identify social-psychological problems, group-specific language configurations, different lexicons, different communication cultures, and communicative insecurity caused by different communication patterns. The first two categories are outright dismissed as being linguistic in nature; the social-psychological problems have to do with what resonates with people and their ideas of the world, and the group-specific language configurations have to do with (differences in) prestige attributed to different dialects and varieties. Lexical problems are obviously linguistic in nature, but do not cause the conflict-like disruptions between East and West German - they might merely cause mutual misunderstanding and alienation. What remains are different communication cultures and communicative insecurity; these, according to Antos & Schubert, are the main reasons for the communicative issues between East and West Germany.
The difference in communication cultures comprises multiple subcategories of differences. First off, the stylistic differences described in 2.3 can be analysed as frequency differences - certain styles appear more on one side than on the other. This means that people choose different styles based on frequency in their language, and therefore cause different communicative patterns in the way they speak (i.e. in their communication culture). Furthermore, both parts of Germany

have different discourse traditions - this influences both style, text type, prosody, and lexicon. East German speakers also have a different communication attitude; this is expressed mainly in different speaking styles and in how people portray themselves in communication. Lastly, there are differences in discourse patterns - preferences, strategies and perspectives may differ between East and West German speakers.
After the Reunification, East Germans became very insecure in their discourse, mainly about public discourse - i.e., towards West Germans. Both sides attributed this to communicative differences between them, but it was mainly the East Germans having problems adjusting to many characteristics of West German discourse. This insecurity manifested itself in longer breaks and delays in speech, as well as in a phenomenon known as hedging - defined by Markkanen & Schröder (1996) as:

> "qualification and toning down of utterances and statements in order to reduce the riskiness of what one says"
>
> "mitigation of what might otherwise seem too forceful"
>
> "politeness and respect to strangers and superiors" (p. 2)

East Germans, in short, adjusted their choice of words and were purposefully more indirect in order to avoid sounding too harsh, strange, or disrespectful to West Germans (Kreutz, 1997). The insecurity about communication among East Germans could be seen as an extension of insecurity about their social identity; unlike Antos & Schubert (1997), Schlosser (1991) sees a direct correlation between that and East Germans' speech. The journal that published Schlosser's article ran a competition in a previous edition, in which they asked their readers to come up with names for the states that were formerly GDR. Whereas West German readers tended to come up with more neutral terms like *Ostdeutschland* ("East Germany") or *Neubundesländer* ("new states"), East German readers were much harsher; Schlosser names, among others, *Aufbauland* ("land under construction"), *Billigländer* ("cheap states"), and *Deutsch-Ostafrika* ("German East Africa"). According to Schlosser, this kind of speech reflects insecurity about East Germans' identity; this may well be the cause for other communication issues.
Several researchers in the last two decades have posited that even though the physical wall dividing Berlin, and by extension Germany, has long disappeared, there is still a wall in people's heads: the *Sprachmauer* ("speech wall"). Plewnia & Rothe (2009), for example, assume that years of being physically separated must have left some traces, and they find that different speech attributes associated with either East or West Germany are still being judged differently by participants from East Germany than they are by West Germans. In addition, East and West Germans continue to have different communicative attitudes towards speech, e.g. in that accents have different prestige for participants from East Germany than they do for West Germans. They do, however, conclude that the bigger picture shows that the way Germans view language doesn't differ much in East and West anymore - the differences are still there, but that *Sprachmauer* is not insurmountable. This view is supported by Kennetz (2010) and Schlobinski (2015); the latter identifies additional sociolinguistic factors like migration patterns and subcultures that will soon have a larger influence on language than the division East-West.

Chapter 3: Literature-based approach

3.1 - Method

The aim of this approach was to ascertain whether lexical differences between East and West Germany found previously and documented in literature, are still present to this day. To achieve this, a subset of all available newspapers from 2019 was extracted from the DeReKo-corpus. This corpus, first created in 1964 and compiled, monitored, and expanded by the *Institut für Deutsche Sprache* (Institute for the German Language, IdS) in Mannheim, encompasses more than 50 billion German written words from various sources like newspapers, magazines, and Wikipedia. The corpus is almost fully accessible online for smaller inquiries via Cosmas II, and new data are continuously being added - generally, all newspapers from the previous year will be accessible around February of the next year.

The newspapers for this approach were selected from DeReKo based on region first and foremost, with the aim to include at least one large newspaper from every *Bundesland*. A total of 19 newspapers were selected; see fig. 1 for all the selected newspapers and their locations projected onto a map of Germany. In this same figure, a division by region has also been added. In order to look at differences between East and West Germany, but also to account for possible variance within those two, we divided the two parts according to the orientation of the newspapers within East or West Germany.

A 'large' newspaper was defined as having at least several million, but ideally several tens of millions of words present in the corpus. This generally meant we took the largest newspaper available for every *Bundesland*, with a total of between 4 million words (*Der Spiegel*, Hamburg) and 63 million (*Rhein-Zeitung*, Rheinland-Pfalz). On average, 24.1 million words were present for each newspaper. In fig. 2, an overview of the amount of words per newspaper is shown.

A second newspaper from Hamburg was selected to compensate for the low word count (*Die Zeit*, 6 million words). A total of three newspapers were selected for Berlin, to account as much as possible for linguistic phenomena caused by it being a metropolis with many international influences. No newspapers were added for Bremen, as none of the ones present in DeReKo even came close to having more than a million words in the corpus. As Bremen is a small Bundesland completely encircled by Niedersachsen, however, it is reasonable to assume that newspapers from Niedersachsen would also be representative for Bremen; it is unlikely that with non-enforced borders and universal access to newspapers and internet, Bremen upholds a fundamentally different language to Niedersachsen. For Brandenburg, no newspaper was included, mainly because any available newspapers were very small. A review of the literature did not indicate that significant linguistic differences were found between Brandenburg and other East German *Bundesländer*; therefore, leaving Brandenburg out was not expected to have any effect on our research.

Fig. 1: Map of Germany showing the division into subregions and the origins of the 19 newspapers

As detailed in the previous chapter, six classes of lexical differences are generally distinguished in literature (cf. Hellmann, 1984): *Lexemspezifika* ("lexem specifics"), *Bedeutungsspezifika* ("meaning specifics"), *Bezeichnungsspezifika* ("designation specifics"), *Gebrauchsspezifika* ("usage specifics"), *Häufigkeitsspezifika* ("prevalence specifics"), and *Wertungsspezifika* ("evaluation specifics"). For the present study, only the *Lexemspezifika, Bezeichnungsspezifika,* and *Häufigkeitsspezifika* were feasible to take into account. The nature of the data makes it impossible to discern potential positive or negative attitudes towards words, which would be necessary for *Gebrauchsspezifika* and *Wertungsspezifika*; furthermore, (subtle) differences in meaning (*Bedeutungsspezifika*) are impossible to detect from newspaper texts alone - a human reader would have to go through all the data to find those, and that is not a feasible task, given the sheer volume of data.

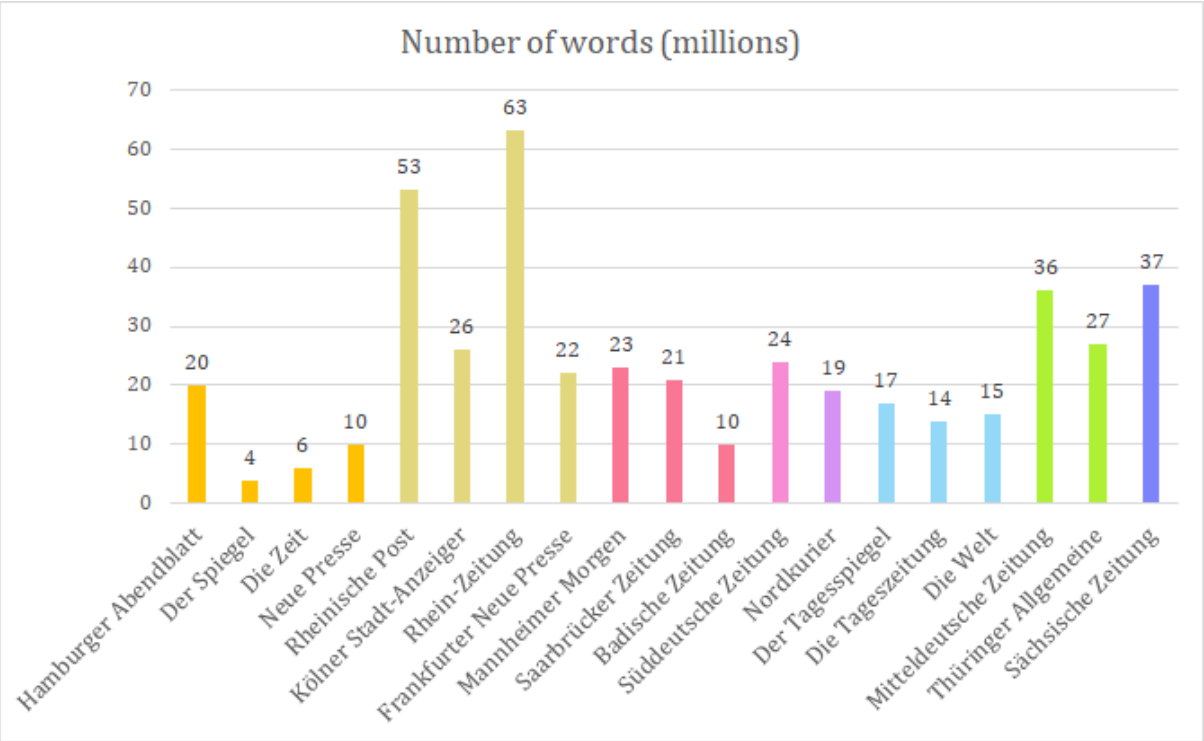

Fig. 2: Amount of words (in millions) included in the dataset per newspaper

A list of words that have previously been designated as *Lexem-, Bezeichnungs-,* or *Häufigkeitsspezifika* was compiled, including their source (see Appendix A). This list is not exhaustive, but contains approximately 80 words and word combinations that were, at one point, definitively specific to either East or West Germany.
After acquiring the desired DeReKo-data from the *Institut für Deutsche Sprache*, samples of 1 million words were made for each newspaper. A word was included in the final data set only if it was present in 5 or more samples, to weed out obscure words. A program written in Perl was run that calculated for each word, bigram, and trigram how often it appeared per million words in every newspaper - mean and standard deviation were given for each newspaper as well as for each region. Bootstrapping was done for all newspapers to ensure there was enough data to work with. This process resulted in a total of 7.8 million unique words, bigrams, and trigrams; because of the nature of the data - individual sentences without further context -, research into the unique words was initially restricted to n-grams. If the results warranted it, for instance to get a (limited) view of the context, trigrams around individual words were later generated.
For each individual N-gram, a bar plot was made using ggplot2 in R, displaying the mean frequency of occurrence per million words in each newspaper. Even though error bars could be generated as the standard deviation was also calculated, the error bars will be left out in the graphs. They display how consistently the word is used throughout time, rather than being an accurate and easily-readable depiction of the actual usage of the word. If warranted and beneficial to the comprehension of the results, two graphs may be displayed next to each other - e.g. for a traditionally East - West German pair of words, or if two graphs together show a phenomenon or tendency.
The *Bezeichnungsspezifika* can additionally be displayed in a different bar chart, if both words are present in the data set and it is likely that those two words are the only alternations available. In that case, a value between 0 and 1 is displayed for every newspaper, where 0

means a newspaper used the East German variant 0% of the time, and 1 means the newspaper used that variant 100% of the time.
The same type of graph was used for words that could feasibly be capitalised because of placement at the beginning of a sentence, e.g. adverbs like *vielleicht* or *etwa*, as well as adjectives and personal pronouns; the graph then shows how often each newspaper capitalised that word. This way, potential (syntactic) capitalisation differences between East and West German newspapers can be visualised.

3.2 - Results

The graphs that were generated in R based on the list of *Häufigkeits-*, *Lexem-*, and *Bezeichnungsspezifika* found in literature, were roughly divisible into four categories.

Cat. I - Words that have fallen into disuse

In total, the data were divided into 485 chunks consisting of ~1 million tokens each; these tokens made up 7.8 million types, including single words, bi-, and trigrams. Of the 77 words included in the list (see Appendix A), only 65 were present in five or more of the 485 chunks of tokens - 12 words (16% of the total) that used to be defining linguistic differences between East and West Germany were used very infrequently or not at all in newspapers in 2019. This included five East German *Bezeichnungsspezifika* like *Werktätiger* ("employee"), or *Kaderleiter* ("personnel manager"), as well as two West German *Bezeichnungsspezifika*, which are all listed in table 1.

Region	**N-gram**	**Meaning**	**Other region's equivalent still present?**
East Germany	Werktätiger	employee	yes (*Arbeitnehmer*)
	Kaderleiter	personnel manager	yes (*Personalchef*)
	Kombine	combine harvester	yes (*Mähdrescher*)
	Popgymnastik	aerobics	no
	nü	yes	yes (*ja*)
West Germany	Aerobik	aerobics	no
	Heimat-vertriebener	refugee	yes (*Umsiedler*) - also West German *Flüchtling*

Table 1: Bezeichnungsspezifika *from the list in Appendix A that are no longer present in the dataset*

In addition, two East German and four West German *Häufigkeits-* or *Lexemspezifika* were found in fewer than five samples of any newspaper and therefore excluded; these are listed in table 2.

Region	N-gram	Meaning
East Germany	umfassender Aufbau	comprehensive structure
	wissenschaftlich-technische Revolution	scientific-technical revolution
West Germany	demokratischer Sozialismus	democratic socialism
	konzentrierter Aktion	concentrated action
	o.k.	okay

Table 2: Häufigkeits- and Lexemspezifika *from the list in Appendix A that are no longer present in the dataset*

Cat. II - Infrequently used words

A majority of the remaining N-grams, 47 in total, are not widely used; for the purposes of this research, "not widely used" is defined as appearing more than 0 times per million words, but less than ten times. As explained in the previous chapter, the standard deviations for these mean values of frequency per million words will not be displayed as error bars. See fig. 3 for an example of how the graph displaying the frequency of *freiheitlich* in fig. 4 would look with these error bars; keep in mind that as a result of the omission of the error bars, there may be errors in measurement that are not immediately visible. Note that all graphs in this section will have their own scale; the y-axis does not display the same values throughout.

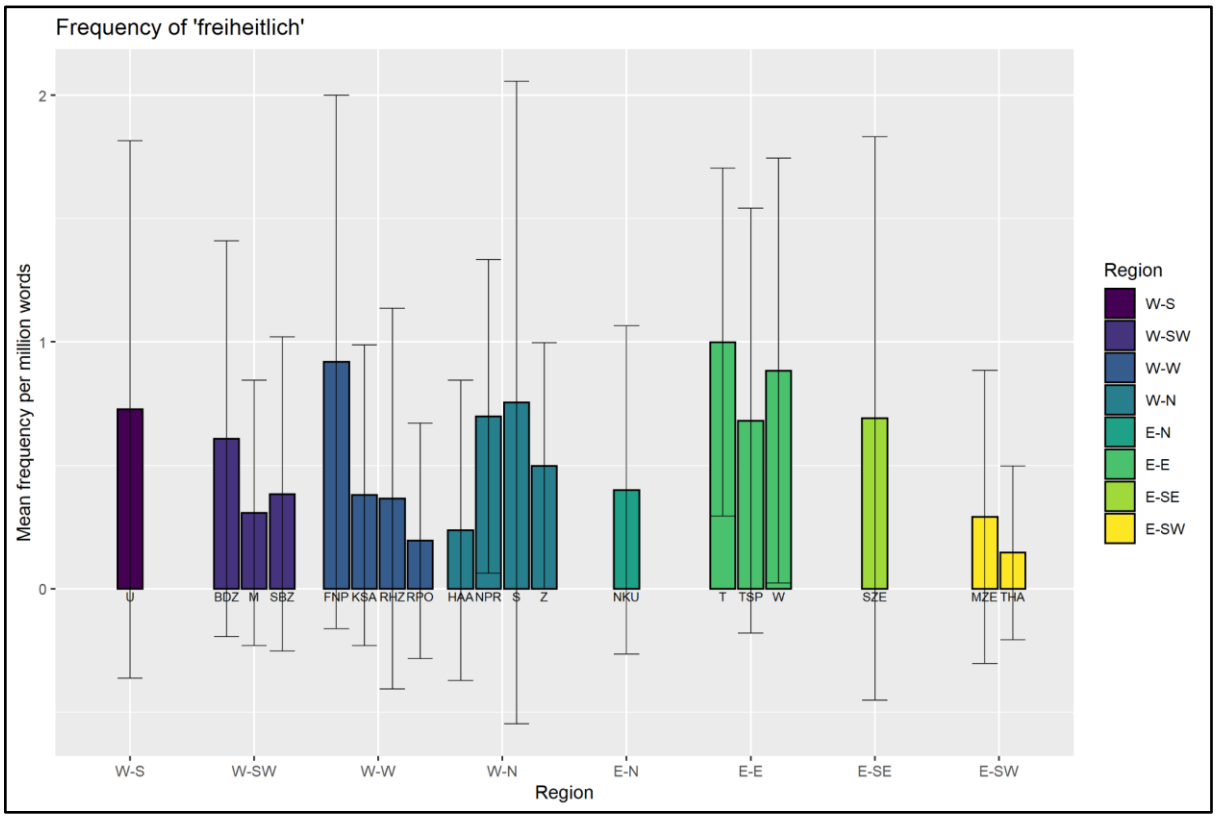

Fig. 3: Mean frequency of appearance per million words of freiheitlich*, displayed with error bars*

First, several words display no significant differences in frequency between East and West Germany. This includes the following *Häufigkeits-* and *Lexemspezifika*: West German *Demokratisierung, freiheitlich* (fig. 4)*, Kids, Prager Frühling;* East German *friedliebend, Massen, Qualifizierung*. One word does: *allseitig*. Traditionally an East German word, it is used slightly more in East Germany to this day.
For *Bezeichnungsspezifika*, the situation gets a bit more complicated, as not all pairs are in the same category. The only pair of words in this category forming a *Bezeichnungsspezifikum* that show no differing tendencies for East and West Germany is *Fruchtsaft* and *Juice*.

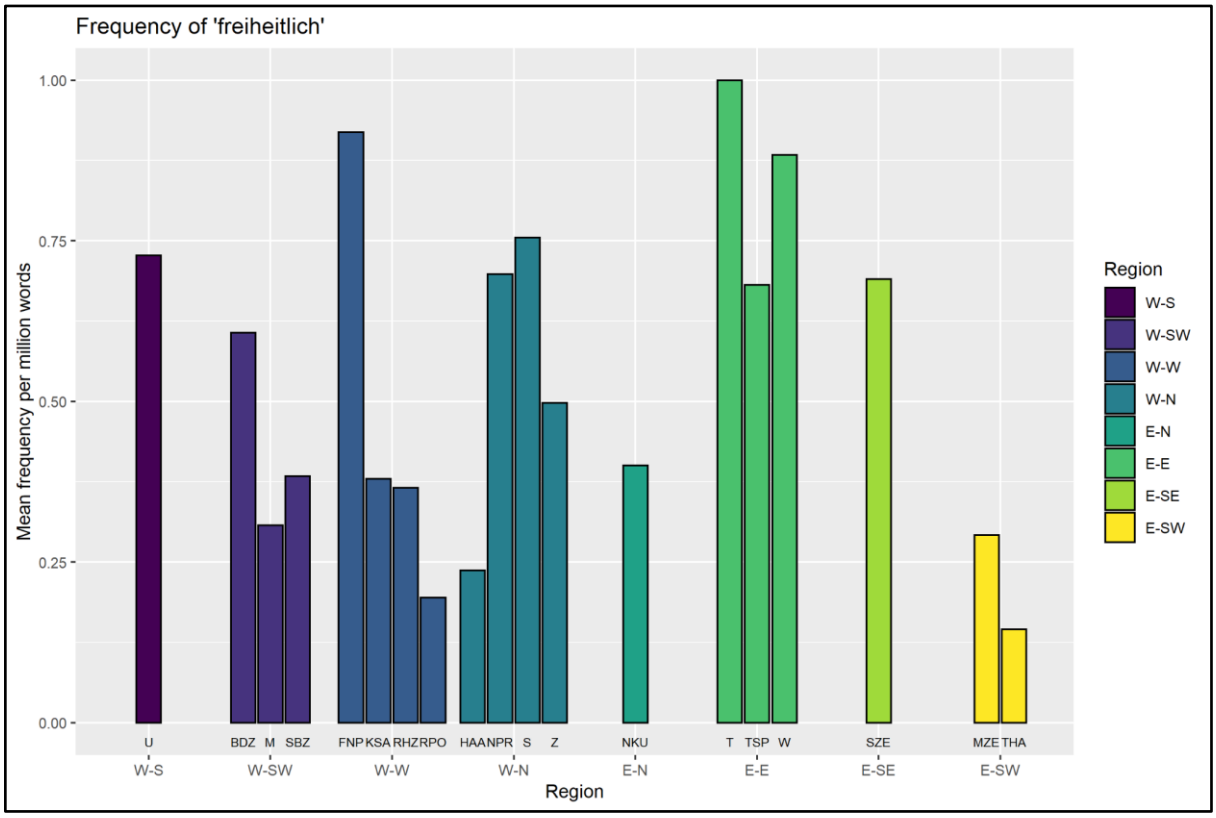

Fig. 4: Mean frequency of appearance per million words of freiheitlich

Although neither of the two words are used in every single newspaper, the *Bezeichnungsspezifika Diskjockey* and *Schallplattenunterhalter* are a good example of one word not being specific to either East or West anymore, while the other is still preferred in one of the two. In this case, *Diskjockey* is used throughout the regions, whereas *Schallplattenunterhalter*, traditionally East German, is still used more in that region (fig. 5). Similar patterns can be found for East German *Kaufhalle* (see fig. 7).

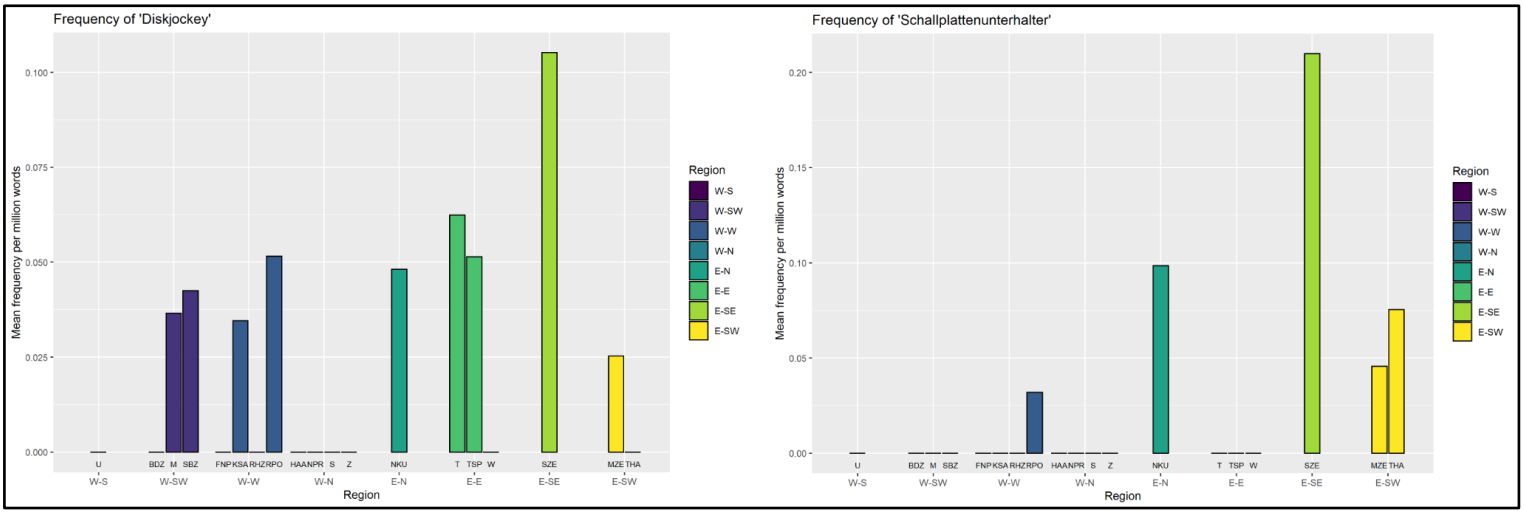

Fig. 5: Mean frequency of appearance per million words of Diskjockey *(left) and* Schallplattenunterhalter *(right)*

Three pairs of *Bezeichnungsspezifika* still display the same usage tendencies as literature has suggested was the case for decades. The clearest examples of this are *Zielsetzung* (traditionally West German) and *Zielstellung* (traditionally East German), as depicted in fig. 6. While keeping in mind that the y-axis does not have the same value for both graphs, there is a clear tendency for *Zielsetzung* (left) to still be used more in West Germany, and for *Zielstellung* (right) to be more common in East Germany.

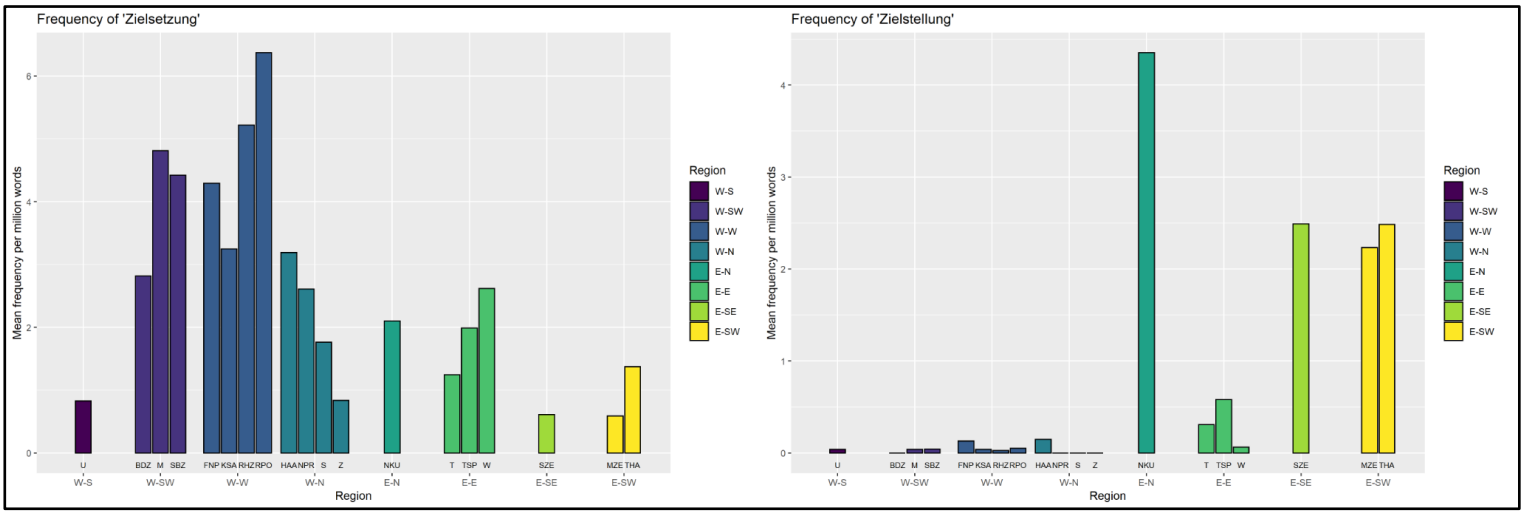

Fig. 6: Mean frequency of appearance per million words of Zielsetzung *(left) and* Zielstellung *(right)*

The same continues to be true for *Brathähnchen* (West) and *Broiler* (East), and *Nachholbedarf* (West) and *Nachholebedarf* (East).

It becomes evident from several graphs that the three newspapers situated in Berlin (region E-E) sometimes collectively display behaviour that differs from what the rest of East Germany does. See for instance fig. 7, in which it is evident that *Kaufhalle* has a higher mean frequency in East German newspapers, but not in the three newspapers from Berlin. In figs. 8 and 9, two

more words (*Demokratisierung*, traditionally West German) and *no* (traditionally East German) show the same phenomenon.

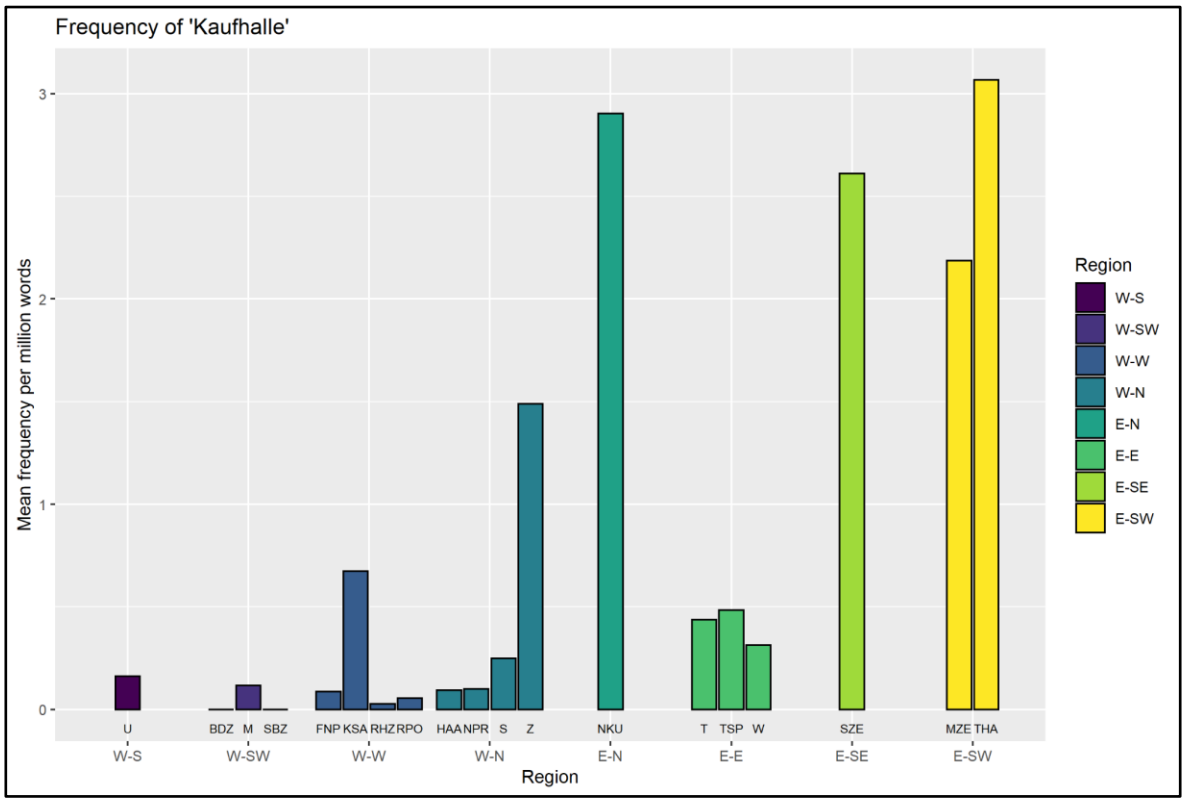

Fig. 7: Mean frequency of appearance per million words of Kaufhalle

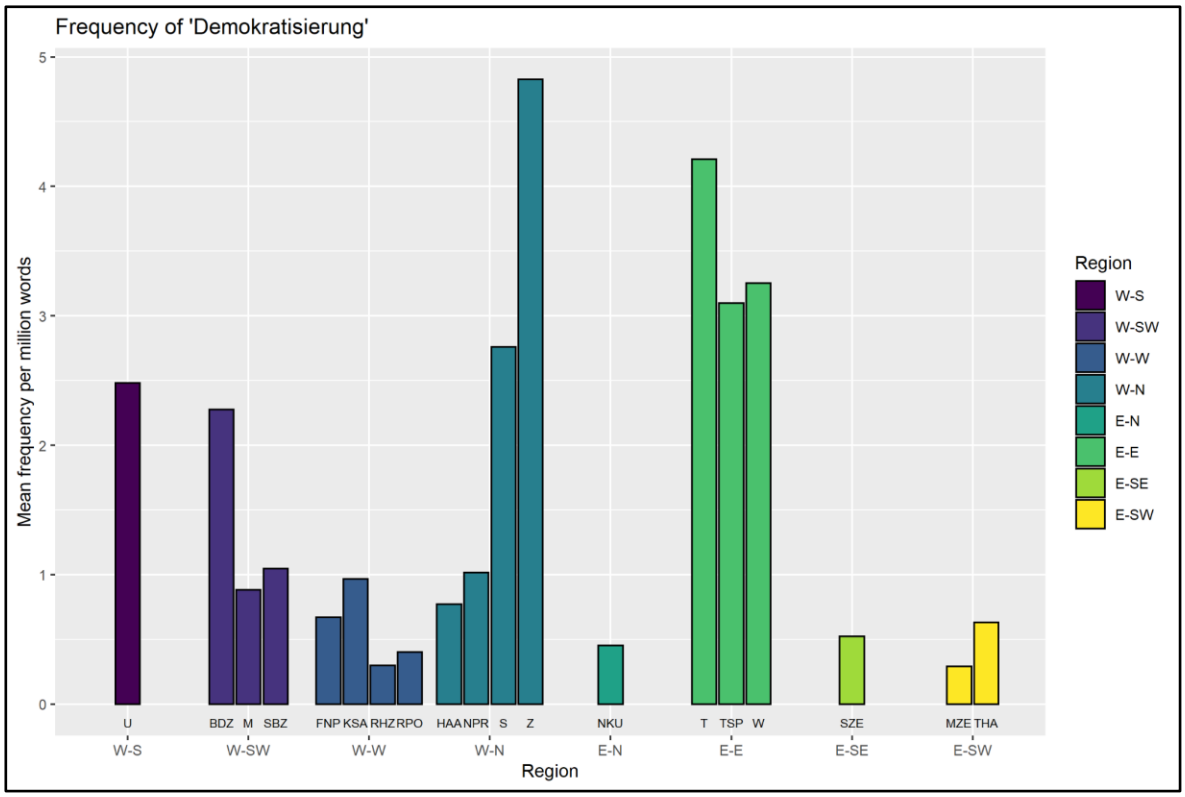

Fig. 8: Mean frequency of appearance per million words of Demokratisierung

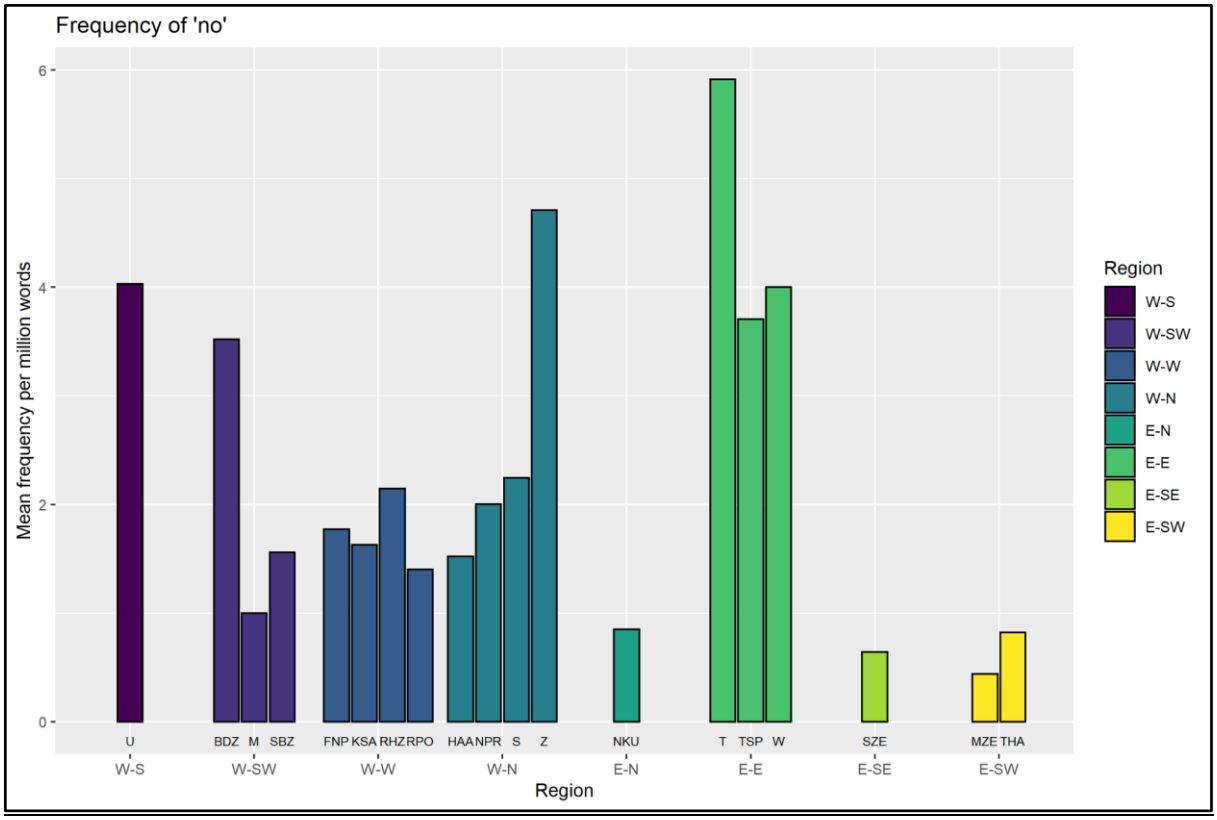

Fig. 9: Mean frequency of appearance per million words of no

Another notable find is that some newspapers display mean frequencies that differ substantially from the other newspapers; see for instance fig. 10 for graphs of the words *Ossi* and *Ossis,* traditionally West German words, which are relatively well used by newspaper "Die Zeit".

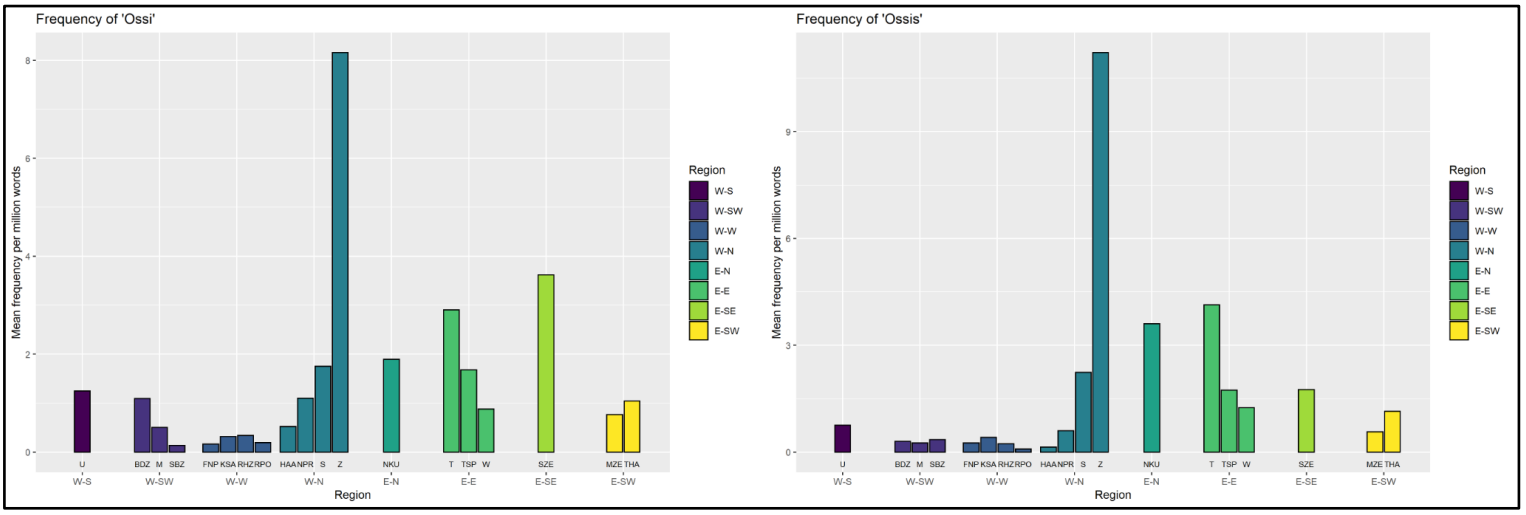

Fig. 10: Mean frequency of appearance per million words of Ossi *(left) and* Ossis *(right)*

Cat. III - Some significant use

This section includes words that appear, on average, more than 10 times per million words, but less than 100 times. There are twelve in total, of which seven are *Häufigkeits-* or *Lexemspezifika* (East German: *Produktion* and *Volk,* West German: *Markt, okay, Partnerschaft, Preis,* and *Super*) and five are part of *Bezeichnungsspezifika.*

The two East German *Häufigkeits-* and *Lexemspezifika Produktion* and *Volk* are displayed in figs. 11 and 12. Neither seem to be favoured in East Germany anymore, with *Volk* even being fairly significantly more frequent in some West German newspapers.

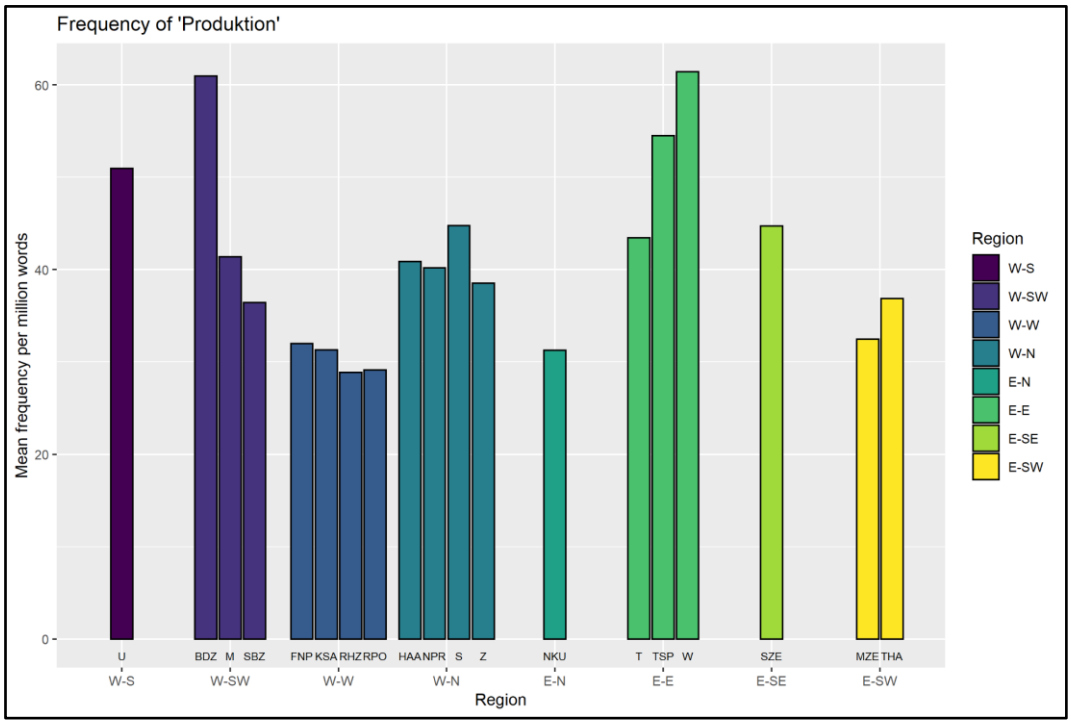

Fig. 11: Mean frequency of appearance per million words of Produktion

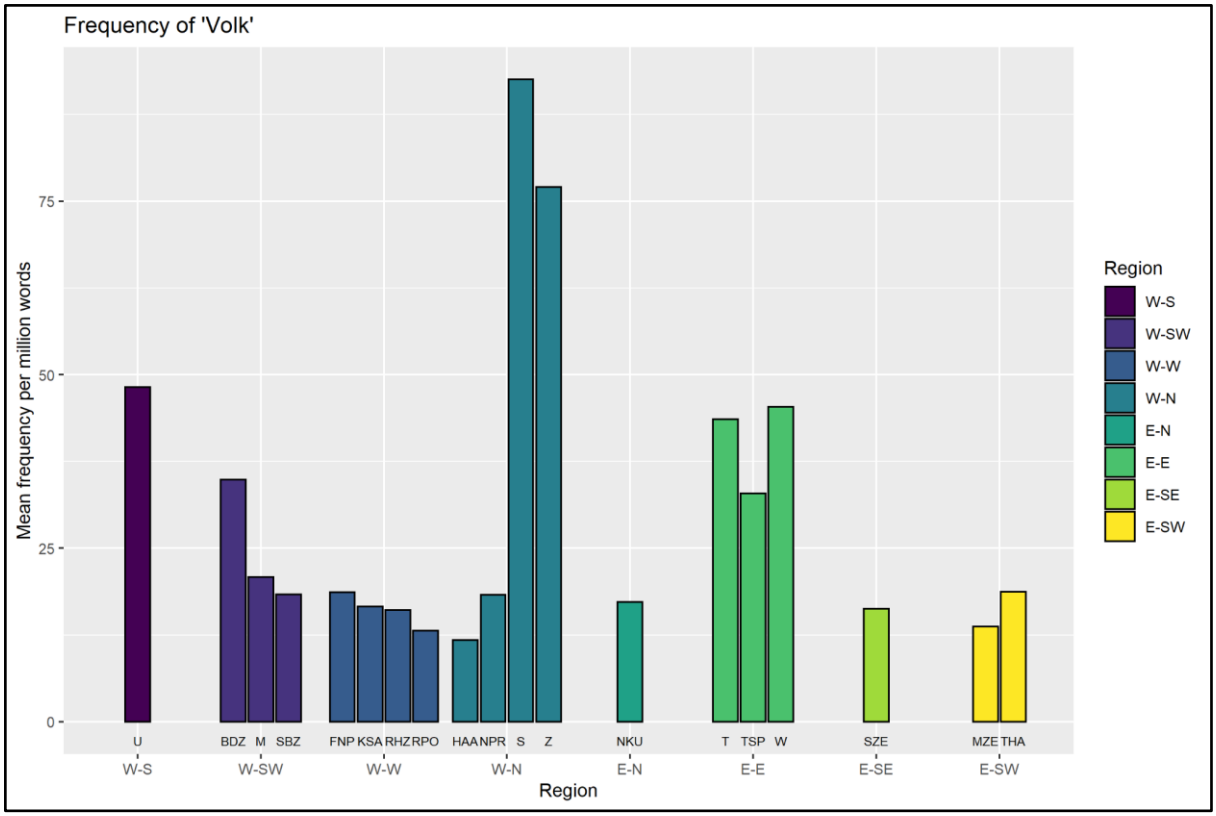

Fig. 12: Mean frequency of appearance per million words of Volk

Of the West German *Spezifika*, there are no significant discernible differences between mean frequency of appearance in East and West Germany. The other two words, however, *Partnerschaft* and *Preis*, display some interesting tendencies; see figs. 13 and 14.

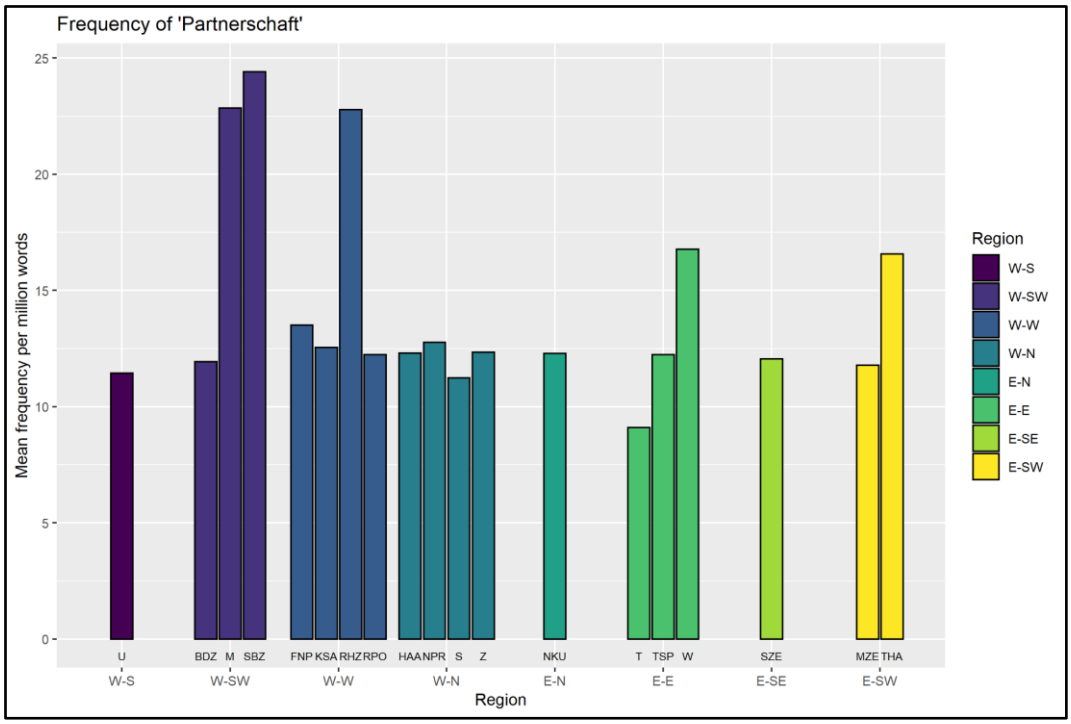

Fig. 13: Mean frequency of appearance per million words of Partnerschaft

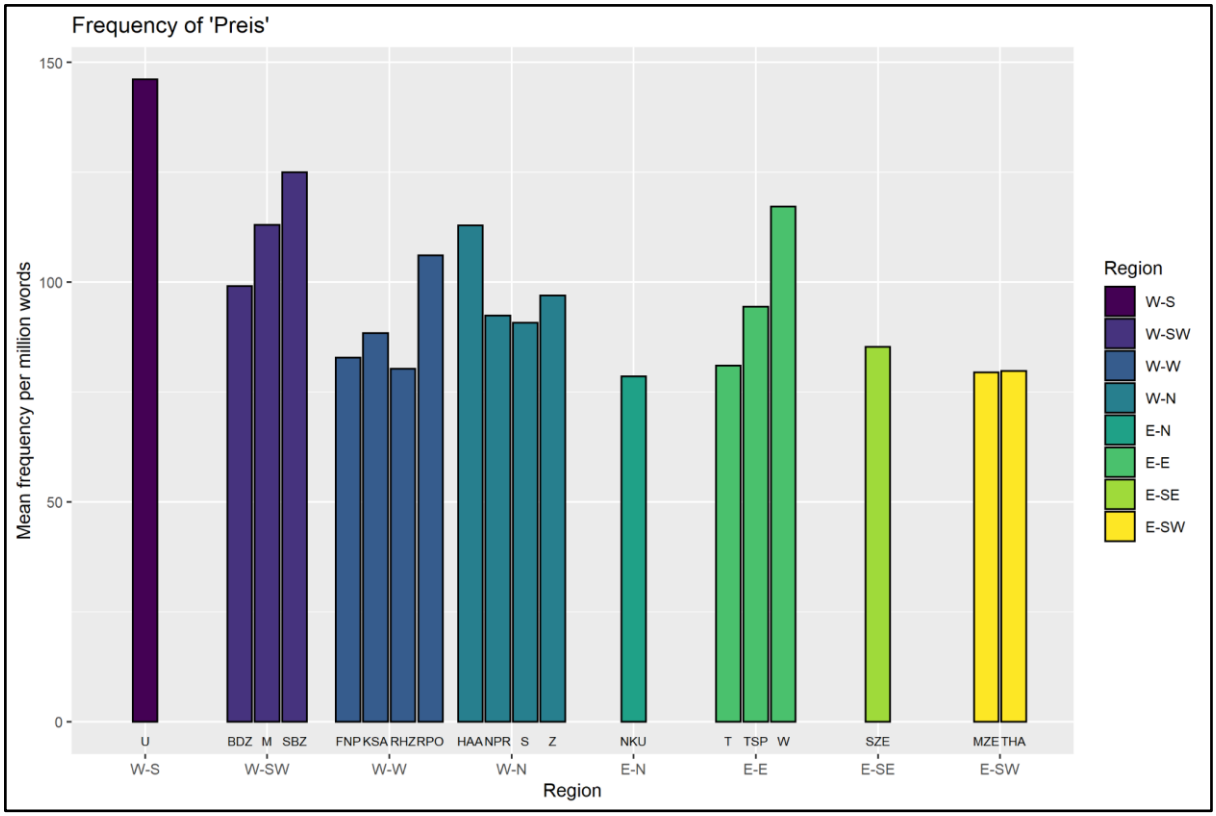

Fig. 14: Mean frequency of appearance per million words of Preis

Overall, *Partnerschaft* appears slightly more frequently in West German newspapers, but this seems to be caused only by three newspapers; otherwise, its frequency is quite uniform. *Preis,* however, seems to still be slightly more prevalent in West German newspapers.

Of the *Bezeichnungsspezifika*, four are West German (*Arbeitnehmer, Kita, Plastik,* and *Supermarkt*) and one is traditionally from the East - *Flugzeug* (see fig. 15). The latter is now used more often in West Germany than in East Germany.
Whereas the traditionally West German *Supermarkt* is used throughout the country nowadays, both *Arbeitnehmer* and *Kita* show abnormal usage in Berlin - *Arbeitnehmer* is used far more frequently by two Berlin-based newspapers than it is in the rest of the country, and *Kita* shows up far less in Berlin newspapers than in most of the rest of the country.

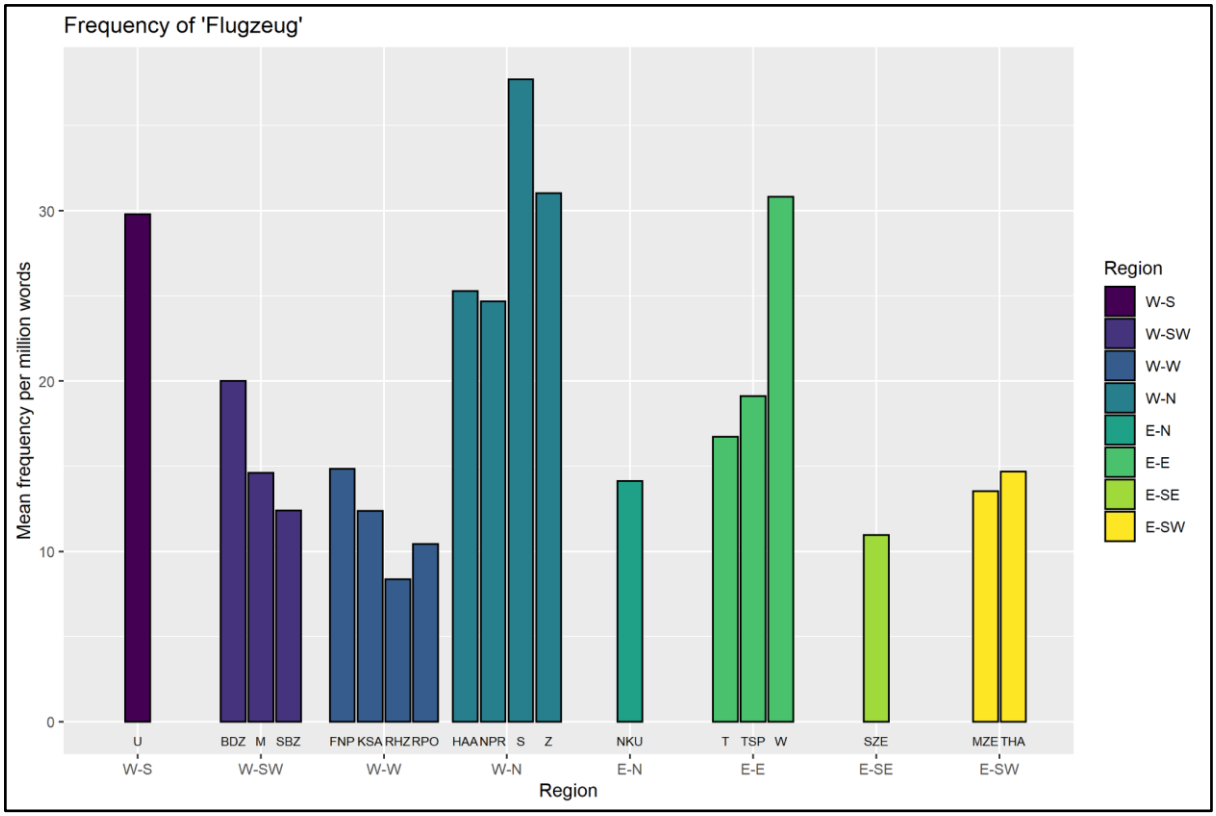

Fig. 15: Mean frequency of appearance per million words of Flugzeug

Cat. IV - Extensive use

Six words appear with a mean frequency greater than 100 words per million. The low number of occupants of this category is to be expected, since most of the words included in this part of the research are not function words. Indeed, four of the words are function words: the pro-sentence *ja* (West German), and the pronouns *ich, wir,* and *unser* (West, East, and East, respectively). The other two are adverbs, *etwa* and *vielleicht* (traditionally both East German). *Unser* and *wir*, historically used more in East Germany, now seem to be distributed equally across both halves (see fig. 16) - note that the y-axis scales are once again not at the same scale.

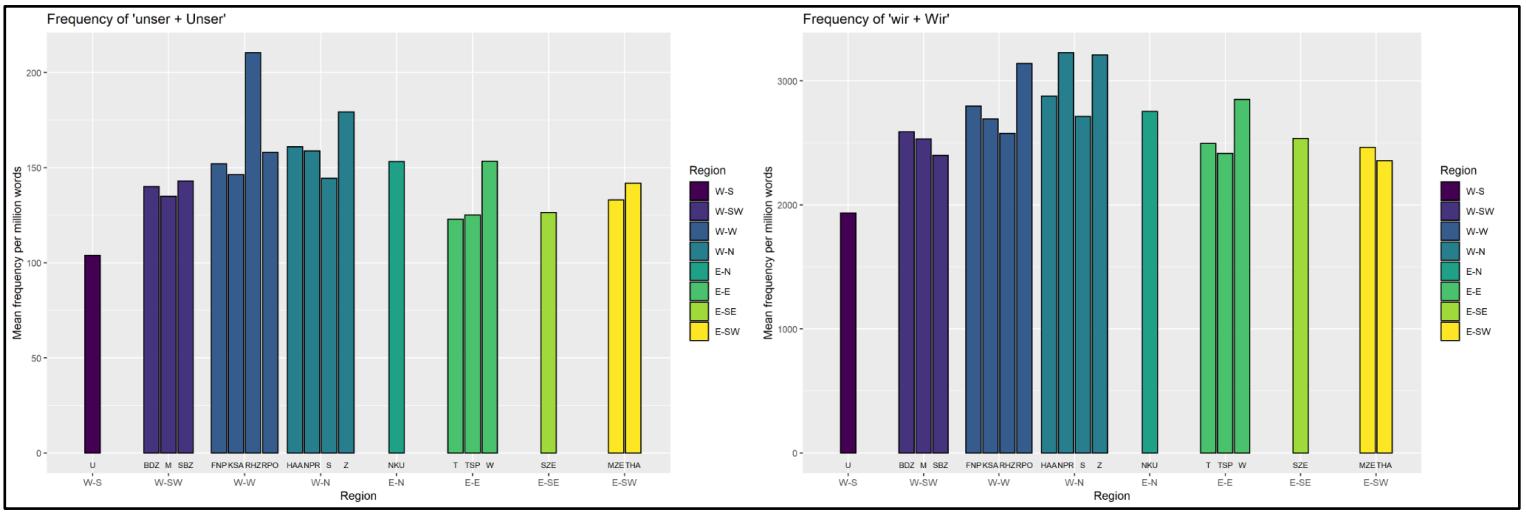

Fig. 16: Combined mean frequency of appearance per million words of the capitalised and uncapitalised Unser *and* unser *(left), and the capitalised and uncapitalised* Wir *and* wir *(right)*

Ich once again displays a peak for a single newspaper (coincidentally, “Die Zeit” again), but seems to otherwise be quite consistent across all newspapers. The same peak, but less extreme, can be seen for *ja* and *vielleicht*, who also seem to be distributed equally otherwise (see figs. 17 and 18). Lastly, *etwa* is relatively homogeneously distributed across all newspapers.

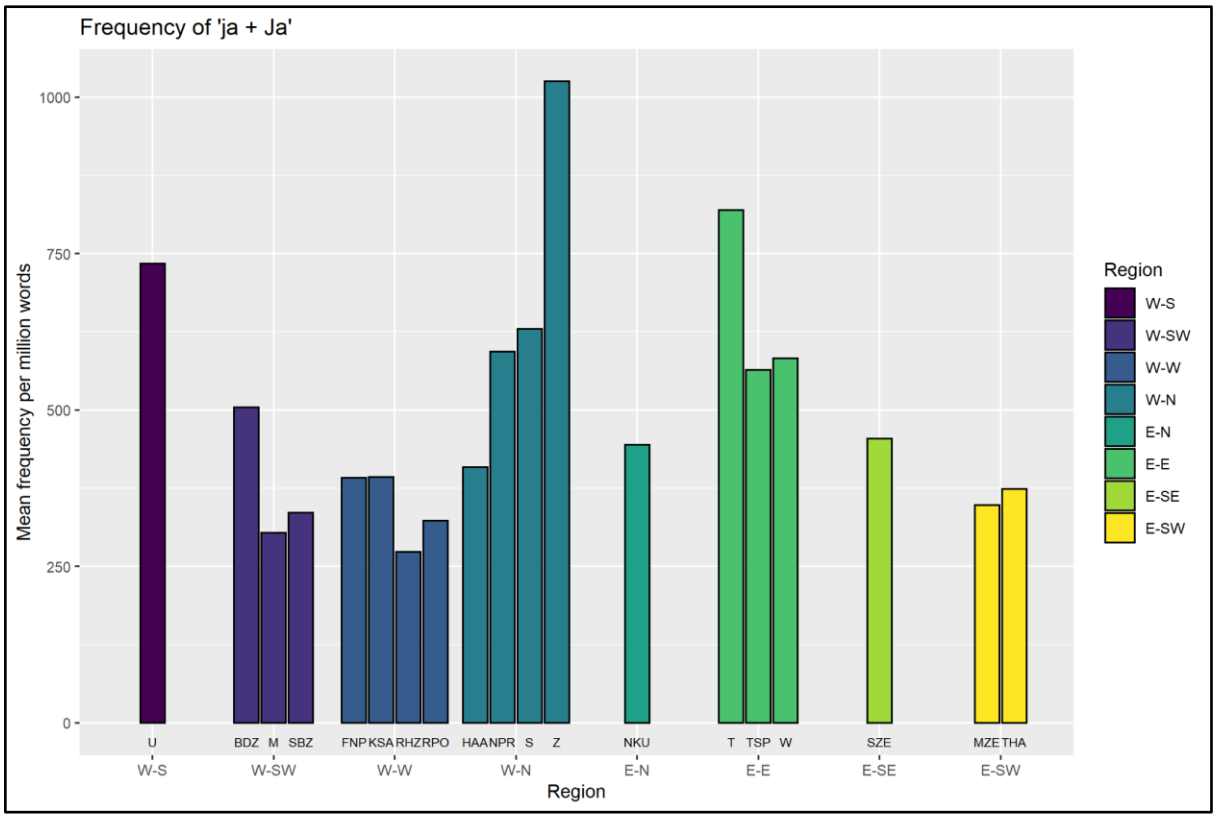

Fig. 17: Combined mean frequency of appearance per million words of capitalised and uncapitalised Ja *and* ja

<u>Capitalisation</u>

A separate section for differences in capitalisation is added at this point, to have a more detailed look at differences between newspapers and regions when it comes to capitalising or not capitalising a word. In total, there were 17 words that had a capitalised version as well as a fully lowercase one. Most graphs did not reveal a pattern of capitalisation that was different between East and West Germany; however, some interesting details are worth mentioning.

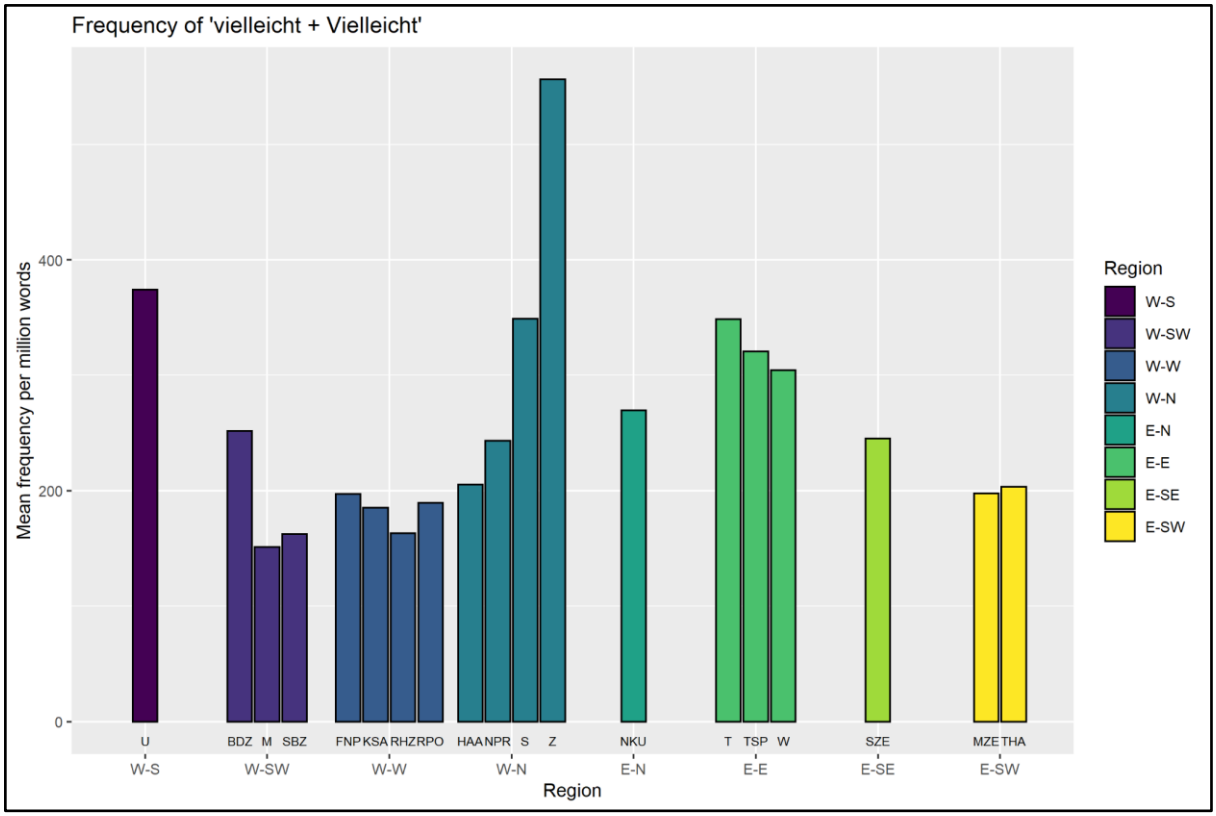

Fig. 18: Combined mean frequency of appearance per million words of capitalised and uncapitalised Vielleicht *and* vielleicht

First, some individual newspapers had a tendency to capitalise certain words to a high degree. The clearest examples of this are *unser*/*Unser* (fig. 19) and *regional*/*Regional* (fig. 20).

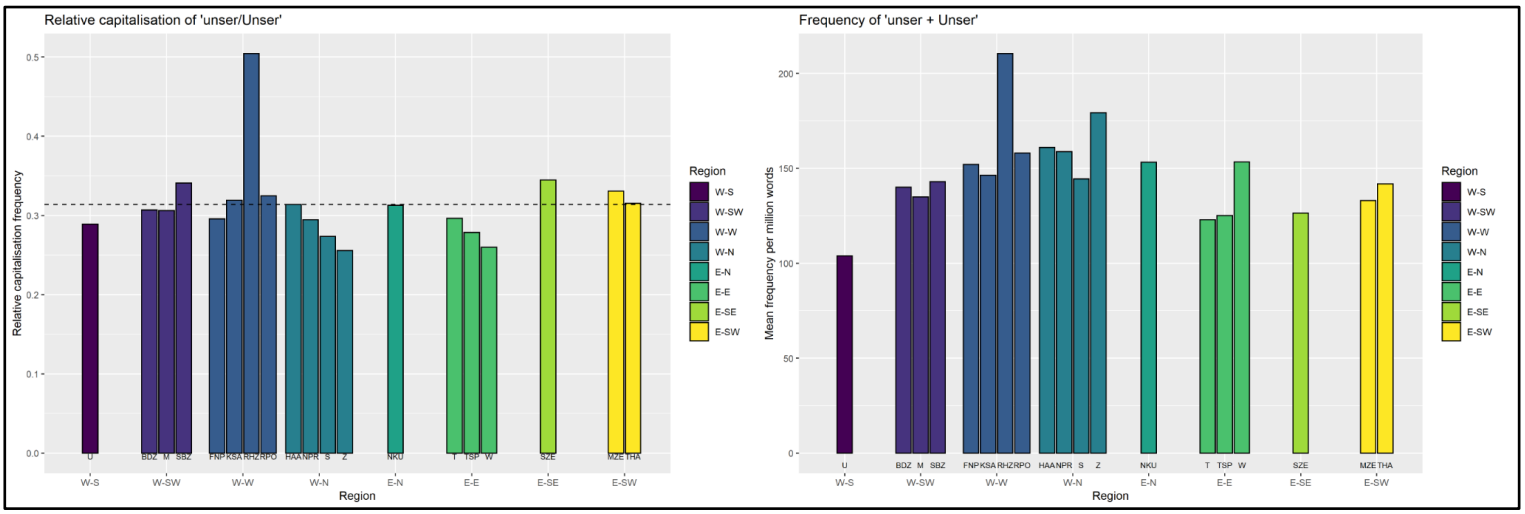

Fig. 19: Relative frequency of capitalisation for Unser *and* unser *(left), and combined mean frequency of appearance per million words for capitalised and uncapitalised* Unser *and* unser *(right)*

Icke and its lowercase version *icke* also showed an interesting dynamic: eleven newspapers printed only the capitalised version. It should, however, be noted that the mean frequency of appearance of both capitalisation versions together did not exceed 1 per million words for any newspaper.

Overall, there were no other significant or surprising differences in capitalisation between East and West German newspapers. Pronouns consistently appeared in a capitalised version between 30 and 40 percent of the time; adverbs either less than 10 percent (*annähernd, ca., etwa*) or about 32% (*vielleicht*) of the time, which is consistent with how logical it is to have that word in the first position of a sentence.

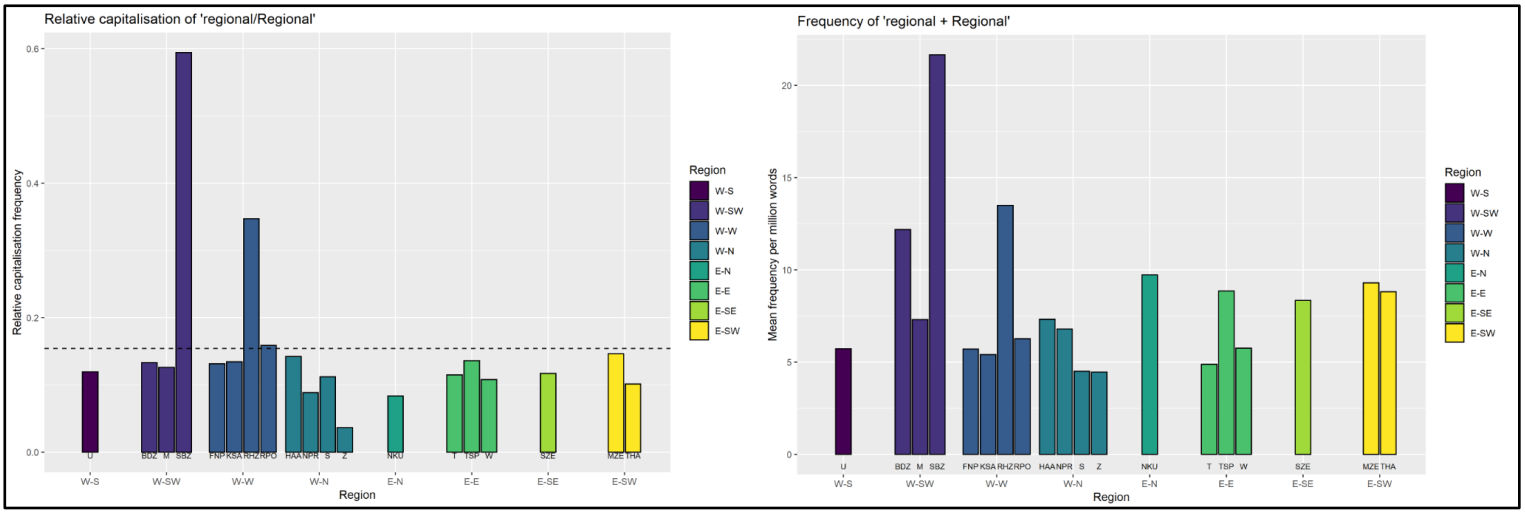

Fig. 20: Relative frequency of capitalisation for Regional *and* regional *(left), and combined mean frequency of appearance per million words for capitalised and uncapitalised* Regional *and* regional *(right)*

3.3 - Discussion

In the first method of this current research, a list of 77 words that were historically noted as differing between East and West Germany (see Appendix A) was compared against nineteen newspapers from all over Germany. Twelve words, or 16% of the total, were too infrequent in the dataset or weren't used at all. This included some bigrams like *demokratischer Sozialismus* or *konzentrierter Aktion*; these were very specific words that were presumably only used in very specific contexts, so their disappearance is perhaps to be expected. Other words like *Werktätiger* or *Kaderleiter* had equivalents in the other part of Germany that were still present in the dataset - *Arbeitnehmer* and *Personalchef*, respectively - and maybe they covered the load better, were less abstract.

Even though the majority of the remaining words were quite infrequent - between zero and ten appearances per million words -, some of their patterns are still the same as they were before 1990. Even though *Supermarkt* is used evenly throughout Germany, *Kaufhalle* is still used more in East Germany. The same pattern appears for *Diskjockey* and *Schallplattenunterhalter*, with the latter (the East German alternative) still showing up more in East German newspapers. Even more so, the pairs *Zielstellung/Zielsetzung* (see fig. 6), *Brathähnchen/Broiler,* and *Nachholbedarf/Nachholebedarf,* with the West German variant on the left and the East German on the right, continue to appear more in their respective regions. There are several caveats to these findings, however. First, the appearance of 'hallmark' East German words like *Broiler* or *Kaufhalle* and, to an extent, their West German counterparts, could be artificially inflated. One quick Google search for *Ost-Deutsche Wörter* leads to many articles about the presence of East German words in today's language (titles along the lines of '*Words that survived the GDR*' or '*46 Words that no [West-German] would understand*'). The nostalgic factor may make it so there are more articles featuring these words. Further, the overall frequency of many of these words is low to very low. For instance, *Schallplattenunterhalter* only appears in five newspapers, of which four were East German (see fig. 5). Even though appearance in such few newspapers is compliant with the prerequisites of our research, it is still important to keep in mind that it may indicate a bias in the data.

Higher frequency words, appearing between ten and hundred times per million words or more than a hundred times per million depending on the category, showed few significant results - the latter category, consisting of four function words and two adverbs, showed no significant differences in appearance between East and West German newspapers. This is not surprising: finding an alternative for function words is harder than for content words - coining a new noun would be easier than adding a new word to the closed word class of pronouns. The same is true, to a lesser extent, for adverbs. Therefore, it is logical that both regions would use the same words equally: a replacement or alternative would not be easily introduced or adopted.

Chapter 4: Research-based approach

4.1 - Method

In Hellmann (1984; summarised in 2008a), corpus-based newspaper research into lexical differences between East and West Germany was already executed. The aim of this approach, therefore, is to provide a replication that is as close as possible to that original research, in order to see which lexical differences are currently present between East and West German newspaper texts .

Hellmann's research was based on the *Bonner Zeitungskorpus* ("Bonn Newspaper Corpus"), a collection of newspapers from East and West Germany with an approximate size of 4.5 million words. Newspapers were added every five years between 1949 and 1974. For West Germany, the newspaper *Die Welt* ("The World") was chosen; for East Germany, *Neues Deutschland* ("New Germany"). The latter was by far the largest and most representative daily newspaper for the East, having the widest reach as well as aligning with the political views of most of the population (i.e. socialist, marxist-leninist). There was no immediate equivalent for the West, so *Die Welt* was chosen, being the newspaper with the widest reach and a political view that aligned with the general view in the west (i.e. moderately liberal-conservative). Although neither newspaper was overtly political, regional newspapers were added for 1964 and 1974 (*Bonner General-Anzeiger,* "Bonn General Gazette" for the West; *Norddeutsche Neueste Nachrichten,* "North-German Latest News" for the East) to lessen the potential impact of covert political views on language, . Furthermore, in 1974, two newspapers were added that were supraregional, but did not necessarily align with any dominant political views: *Frankfurter Rundschau* ("Frankfurt Look") was added for the West, *Der Morgen* ("The Morning") for the East. See fig. 21 for a brief structural overview (Hellmann, 2008a: 260), with an explanation for the abbreviations in fig. 22 (Hellmann, 2008a: 261); *images omitted due to copyright issues*.

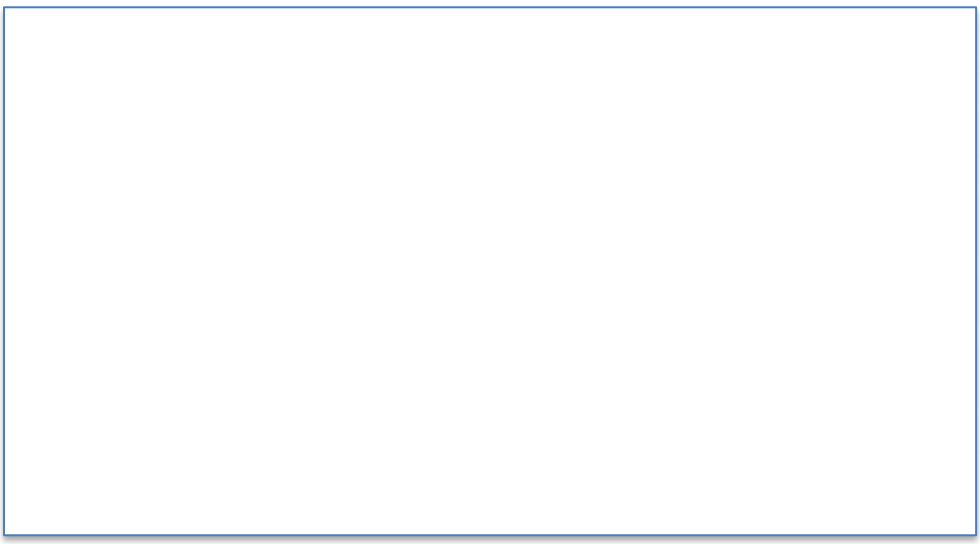

Fig. 21: Composition of the dataset used in Hellmann (1984, 2008a)

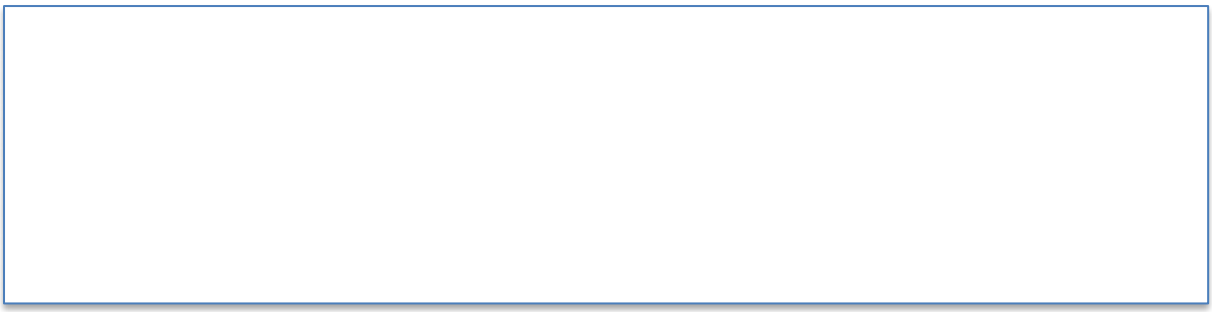

Fig. 22: Explanation of the abbreviations used in Hellmann (1984, 2008a)

The texts from the *Bonner Zeitungskorpus* were split into single words and subsequently analysed by machines, with the goal of researching relative and absolute frequencies. In addition, the aim was to provide the possibility of doing comparative research over several dimensions, such as comparisons between texts from East and West Germany, but also chronological or regional/supraregional comparisons.
For the present study, this approach could not be directly copied. The main reason for this is that the newspapers in the *Bonner Zeitungskorpus* no longer satisfy the criteria with which they were selected. *Neues Deutschland*, for example, no longer has a representative audience in East Germany. Additionally, several of the newspapers chosen in Hellmann (1984) were not included in DeReKo or were not present for 2019. Therefore, a number of adjustments had to be made to the selection process for the newspapers, which were ultimately chosen from the list of nineteen that was compiled for the first approach as detailed in the previous section; the result was as follows.
For East Germany, we found three newspapers that were aligned with the dominant GDR-party SED at some point in time, much like *Neues Deutschland* was. These were *Mitteldeutsche Zeitung, Nordkurier,* and *Thüringer Allgemeine*. Together, they account for a sizable portion of East Germany, which makes them as close as we could get to the criterion of representativeness; *Nordkurier* also has a similar distribution area to what the Eastern regional newspaper had in Hellmann (1984).
For West Germany, two newspapers were found that were aligned with the general political views in the West: *Süddeutsche Zeitung* and *Die Zeit*, from the south and the north, respectively. Both have a sizable distribution area; this is as close as we could get to the representativeness criterion as used in Hellmann (1984). To include some of the midwest as well as have a newspaper comparable to the Western regional newspaper added in Hellmann (1984), *Kölner Stadt-Anzeiger* and *Rheinische Post* were added.

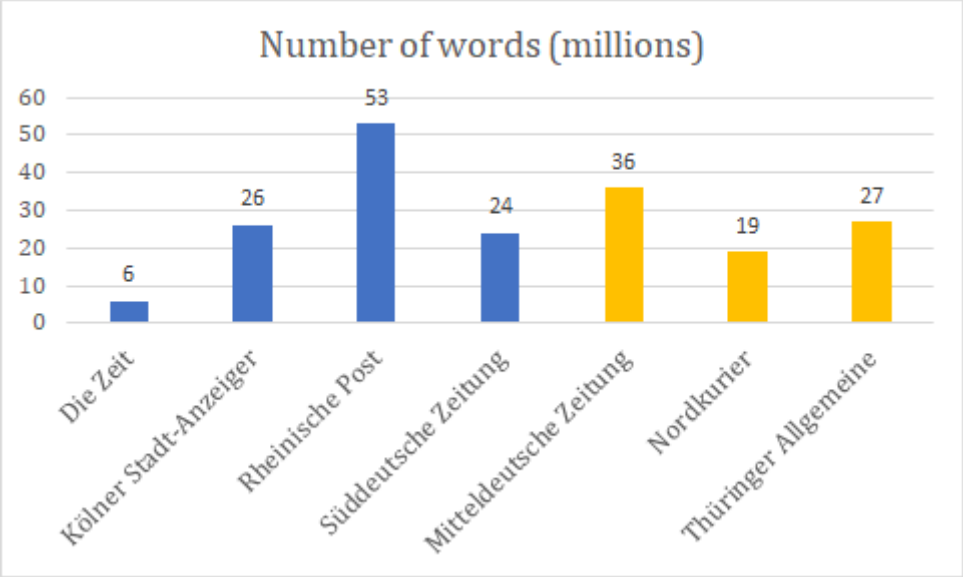

Fig. 23: Amount of words (in millions) included in the dataset for the research-based method

The seven chosen newspapers were extracted from DeReKo; see fig. 23 for an overview of how many words were included per newspaper. Instead of looking specifically for words from a previously compiled list that are known to have been at least somewhat unique to either part of Germany (Appendix A), a statistical analysis was done following the procedures as documented in Hellmann (1984). First, a frequency of appearance per 240,000 words was calculated for every n-gram - 240,000 is the number of words that Hellmann (1984) had at their disposal per region. Seeing as four West German and three East German newspapers were selected for this research, this meant calculating the appearance per 60,000 words for every West German newspaper and the appearance per 80,000 words for every East German newspaper, and then adding all the values for the East and for the West. Subsequently, the same procedure for statistical analysis as done in Hellmann (1984, pp. 235-236) was followed: a p-value was computed with the formula

$$p = \frac{N1p1 + N2p2}{N1 + N2}, \text{ with}$$

p1 the mean frequency of appearance in the East German newspapers
p2 the mean frequency of appearance in the West German newspapers
N1 total amount of words from East German newspapers (here: 240,000)
N2 total amount of words from West German newspapers (here: 240,000)

Afterwards, a z-score was calculated; a value $\geq |1.96|$ indicated a significant difference in appearance between East and West Germany. The z-score was computed with the formula

$$z = \frac{p1-p2}{\sqrt{pq\frac{N1+N2}{N1N2}}}, \text{ with}$$

q = 1-p

Additionally, it was required that N1p1, N2p2 >10.

4.2 - Results

Of 7.8 million unique N-grams, only 3,336 satisfied the criterion of N1p1, N2p2 >10. Fig. 24 shows an overview of all z-scores found, with +5 and -5 as limits. For 102 N-grams, z-scores higher than 5 or lower than -5 were found - these are not shown in the graph below, but are still accounted for in table 3.

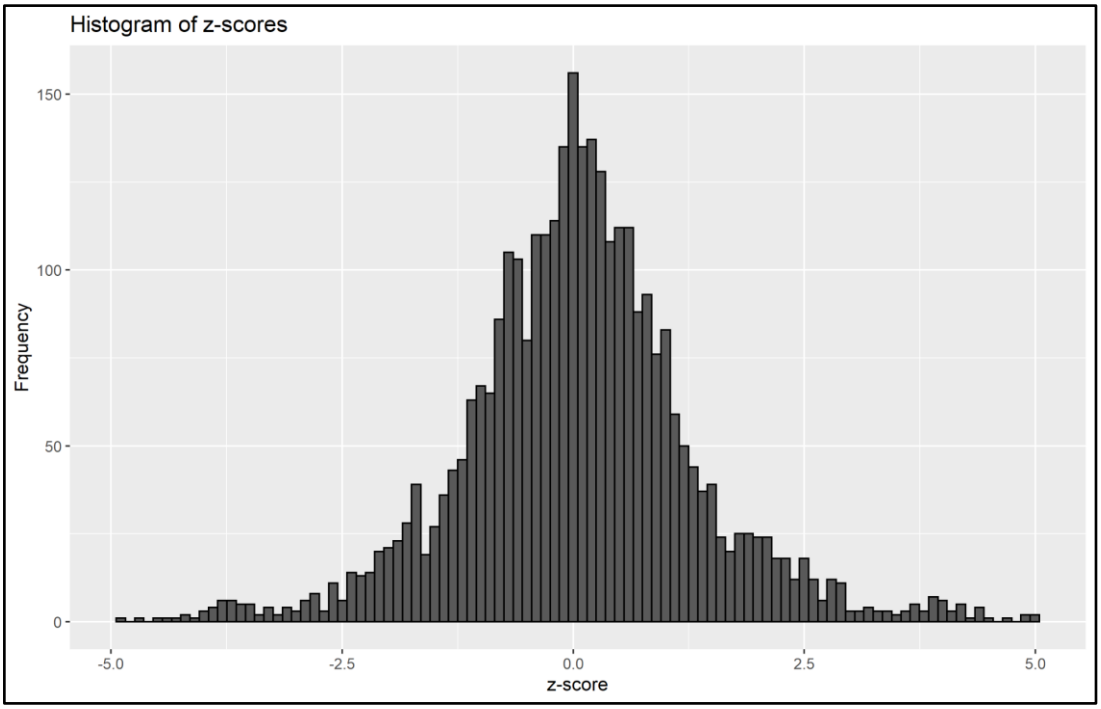

Fig. 24: Distribution of z-scores for the research-based method

A closer look will now be taken at the first and last 40 words - table 3 displays the ones with the highest and lowest z-scores next to each other. A negative z-score indicates a word is used significantly more often in West German newspapers; a positive z-score indicates the same for East German newspapers. N-grams consisting of metadata, quotation marks, or proper names were left out.

N-gram	**Translation**	**z-score**	**N-gram**	**Translation**	**z-score**
,	,	-15.0828	/	/	13.0793
ich	I	-7.86731	am	at/on the (+dat)	9.36957
, und	, and	-7.49809	Uhr	hours	7.69110
sie	she	-7.07195	zum	to the (+dat)	5.42517
:	:	-7.04923	sowie	as well as	5.08707
er	he	-5.89933	der	the (nom sg m), of/to the (gen/dat sg f), of the (gen pl)	5.08338

;	;	-5.51035	Polizei	Police	5.02903
man	one (neutral pronoun)	-5.31955	in der	in the	4.91050
als	as	-5.25345	werden	to become	4.65441
wie	like	-5.23616	Bürgermeister	mayor	4.25884
Sie	you (formal)	-5.22366	, so	, thus	4.19608
Ich	I	-5.08237	Region	region	4.16069
, die	, that/, she	-4.91190	, sagte	, said	4.02698
nicht	not	-4.67114	worden	has become	4.01216
, sie	, she	-4.49902	im	in (+dat)	3.98944
, er	, he	-4.43988	diesem Jahr	this year	3.88342
Welt	world	-4.22400	des	of the (+gen)	3.85633
ist	is	-4.15098	ab	from	3.78414
was	what	-4.14156	14 Uhr	2 PM	3.77469
weil	because	-3.98636	in	in	3.74069
, das	, that/, the	-3.98497	Gemeinde	municipality	3.68633
, als	, when/, as	-3.96144	Jahr	year	3.67453
dass	that	-3.90881	der Region	the region (+gen/dat)	3.62795
wenn	when	-3.84731	in diesem Jahr	in this year	3.58748
Politik	politics	-3.82706	die Polizei	the police	3.55789
, was	, which	-3.79141	sagte	said (1/3sg)	3.45181
, aber	, but	-3.76820	-	-	3.38754
Regierung	government	-3.75013	Feuerwehr	fire brigade	3.38628
hat	has	-3.72454	Euro	euro	3.37949
?	?	-3.65251	wurde	became (1/3sg)	3.32431
mir	to me (1sg dat)	-3.62972	Geburtstag	birthday	3.29806
, dass	, that	-3.62752	vom	from the (+gen)	3.26237
mich	me (1sg acc)	-3.58036	, erklärt	, explains	3.22402

Aber	But	-3.53191	Uhr in	hours at	3.17733
Von	From	-3.50292	sei	would be	3.15417
, in	, in	-3.40276	bis	until	3.14888
ihn	him (3sg m acc)	-3.37604	Am	At/On the (+dat)	3.11742
, weil	, because	-3.31992	Gäste	guests	2.96683
, wenn	, when	-3.31465	Verein	association	2.94335
du	you (sg)	-3.27753	Freitag	Friday	2.94062

Table 3: Top 40 words appearing significantly more in West German newspapers (left) and East German newspapers (right)

The most striking pattern in the West German N-grams is the comma followed by a pronoun or conjunction: there are six of those present in the first 25 N-grams that appear more in West German newspapers, while there are only two in the East German top 25. Expansion to 40 words exacerbates the pattern even more: an additional six N-grams follow this pattern in the words appearing more in West German newspapers, whereas only one is added for the East German newspapers. The latter, ", erklärt", doesn't even follow the pattern - it is a comma followed by a verb. Funnily enough, the most prominent West German N-gram following this pattern is ", und", which is technically grammatically incorrect - a coordinating conjunction should not be preceded by a comma, according to German prescriptive grammar. Several conjunctions as well as some other pronouns appear on their own in the list as well; interestingly, this includes *ich* and *Ich*, *du* and formal *Sie*, but also the neutral pronoun *man.* Both *Politik* ("politics") and *Regierung* ("government") appear more in the West German newspapers. Other interesting features include the more frequent appearance of commas, semicolons, and colons in West German newspapers.

In East German newspapers, we find several forms of the verb *werden* - an auxiliary verb meaning "to become", indicating a passive sentence. Staying with the verbs, *sagte* also appears twice in the list - once following a comma. Furthermore, the police and the fire brigade are mentioned more often in East German newspapers, as well as the phrase *in diesem Jahr* and its components *in, Jahr,* and *diesem Jahr*. In total, seven prepositions (*am/Am, zum, in/in der, im, ab, vom* and *bis*) and two articles with case marking (*der* and *des*) appear more frequently in East German sentences than in West German ones. Four of these are compounds with a dative aspect, one has a genitive aspect.

4.3 - Discussion

For method two, a replication of the statistical analysis of East and West German newspapers as done by Hellmann (1984) was attempted. For seven newspapers, the frequency of n-grams per 240,000 words was calculated. From that, a p-value and a z-score were deduced - a z-score $\geq |1.96|$ indicated a significant difference in frequency of appearance between East and West German newspapers. After manually filtering out noise, such as n-grams that contained metadata or quotation marks, a top 40 of words with the highest z-scores was compiled for both regions. For the full list of words, please refer back to table 3 in the Results section.

First of all, the data seem to indicate that colons appear significantly more in West German newspapers; this should be taken with a (large) grain of salt. Left out were the N-grams *Foto:* and *Foto*, originally occupying the 3rd and 7th position for West Germany respectively, which can be found in captions under pictures. These will have accounted for a significant amount of colons, seeing as *Foto:* in particular had a z-score of -7.55751.

As mentioned in the results in paragraph 4.2, a striking pattern was found in West German newspapers. Along with the comma being the N-gram with the lowest z-score (-15.1) - thus appearing significantly more in West German newspapers -, twelve N-grams that appeared more in West German newspapers consisted of a comma followed by a pronoun or conjunction. This seems to point to a tendency of making longer sentences. The most frequent one was ", und", which is grammatically incorrect - it seems that the pattern of connecting more sentences is preferred even when technically wrong. East German newspapers only saw a more frequent appearance of ", so", ", sagte", and ", erklärt". These are all used to elaborate on a quote, e.g. to mention who said it, and therefore do not point to a tendency to elongate sentences.

An interesting pattern appearing in East German newspapers is the frequent usage of indications of time. Right at the top, we find *Uhr*, mostly used in conjunction as in *14 Uhr*, which is also found on the list. A closer look at the data reveals that this is mostly caused by a select few newspapers: the highest appearance of *Uhr* was in the *Rheinische Post*, whereas the lowest was in the *Zeit* (see fig. 25). Now, both of these are West German newspapers, but quite a few East German newspapers followed the *Rheinische Post* closely in numbers. This could mean that East German newspapers tend to print a TV guide, elaborated on (cultural) events on a certain day, or otherwise mentioned specific times whereas half the West German newspapers generally don't. The pattern of frequent appearance only in certain newspapers seems robust; when looking at *14 Uhr*, as shown in fig. 26, it is again the *Süddeutsche Zeitung* and *Zeit* that barely contain that N-gram. Note that the y-axis scales in figs. 25 and 26 are not the same.

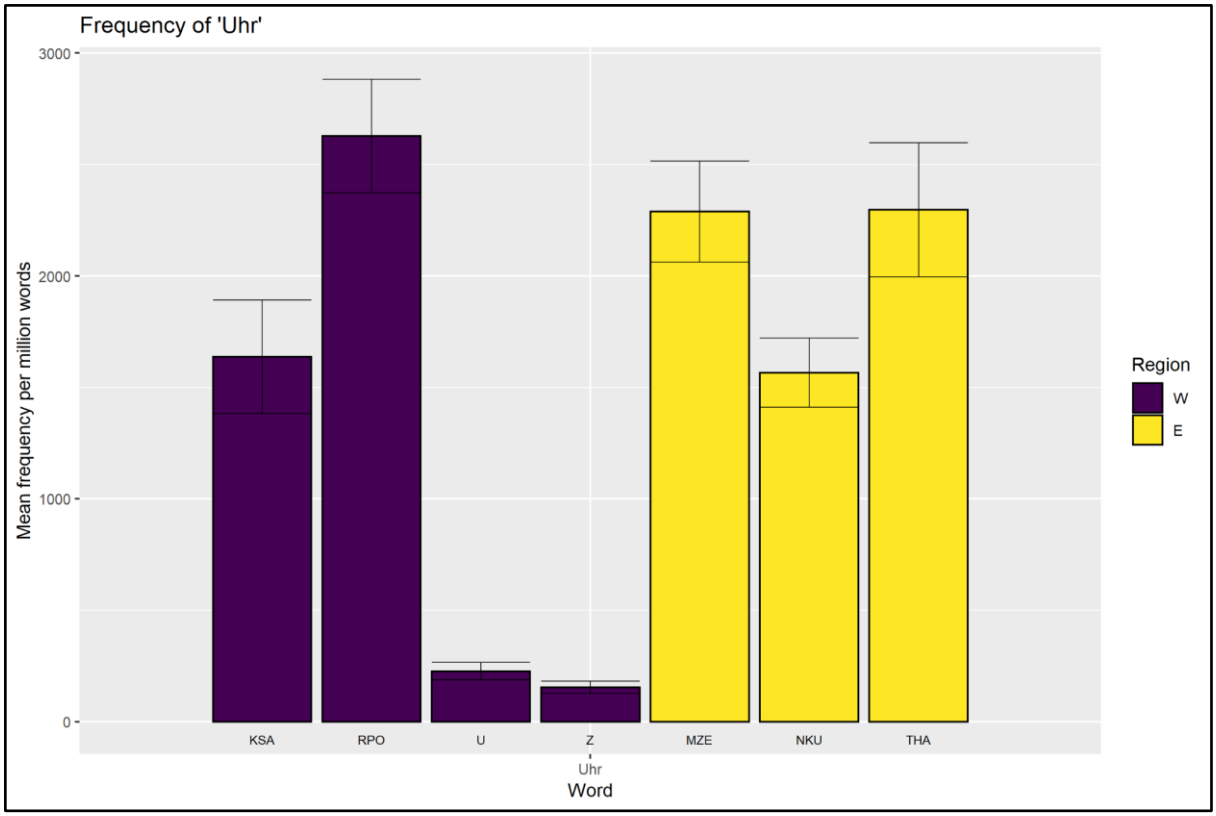

Fig. 25: Mean frequency of appearance per million words of Uhr

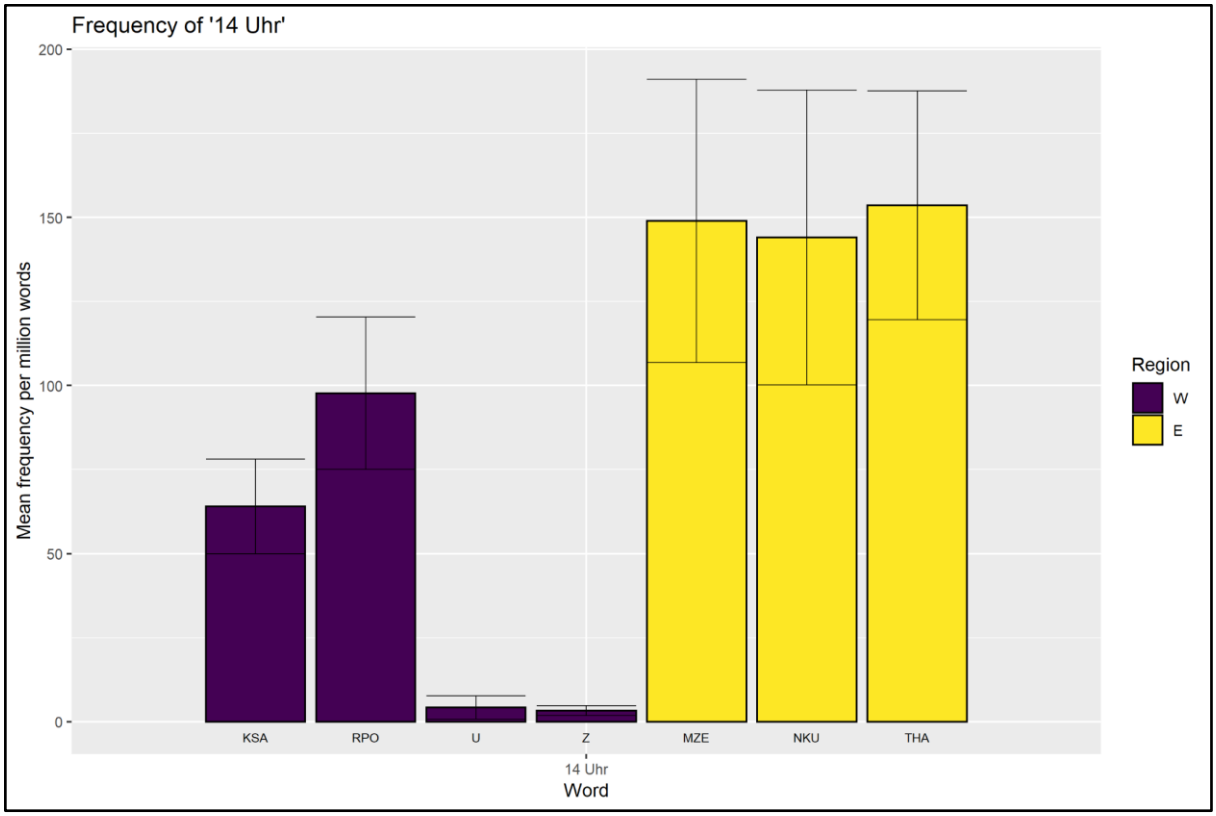

Fig. 26: Mean frequency of appearance per million words of 14 Uhr

Another East German pattern is the presence of several prepositions, most with case marking. There are seven prepositions in total: *am* (and capitalised *Am*, which was not counted again), *zum, im, ab, in* and specifically *in der, vom,* and *bis.* Of the case markings, which in this scenario means they are a contraction consisting of a preposition and an article, four are dative (*am, zum, im*, and *vom*). The other three (*ab, in,* and *bis*) do not have explicit case markings, but they do take a specific case: *ab* generally takes dative, *bis* takes accusative, and *in* can take either. In comparison, there is only one preposition that appears more in West German newspapers (*Von*) and it does not have case marking, although the word after it must be dative. This points to a tendency to use more prepositions in East German newspaper articles - the purpose of a proposition is to express the relationship between a noun and the rest of the phrase, but there are many possible consequences to having more prepositions in a sentence. Without further context, it is impossible to guess the impact this has on the language in East German newspapers.
Lastly, the presence of three different conjugations of the verb *werden* (to become) in the East German top 40 seems to indicate that East German newspapers tend to use more passive sentences. For one of the verb conjugations, *worden*, this is certainly an indicator of the passive; if it were an active verb, it would be *geworden*. The other two conjugations *werden* and *wurde*, however, can be both the main verb in an active phrase or the auxiliary verb in a passive sentence. In case it is the latter, the sentence must further contain a participle (*wurde gesucht*, *verurteilt werden*). A look through trigrams containing any of the three relevant conjugations of *werden* reveals that in many cases, there is indeed a participle close by; however, in many other cases, the only things present in the trigrams are non-participle verbs or other words (*werden kann*, *wurde in den*, *und wurde von*, etc.). Although it is plausible that many of these are also parts of passive sentences, it is impossible to say for certain given the limited context - the participles may well be far away in the sentence. To confirm that East German newspapers do indeed use more passive sentences, further research in which more context is present would be necessary.

Most of the differences elaborated on so far are syntactic differences, or at least not clear cut lexical differences. As described in Chapter 2, lexical differences were usually found in nouns. For West German newspapers, the top 40 N-grams included only three nouns - *Welt*, *Politik,* and *Regierung*. It is funny that two of these appear to be politically inclined: for one, the German government resides in Berlin in the East. At first sight, it is plausible that the West German newspapers are more world- and politics-oriented; the *Rheinische Post,* for instance, has a physically separate section for national and international political news whereas the other newspapers do not appear to have one. However, a look at the data reveals that *Politik* and N-grams containing the word generally seem to be more frequent in the *Zeit* and *Süddeutsche Zeitung*; the reasons for this are not immediately clear.
The East German top 40, then, contains more nouns. These seem to be without pattern, with only (*die*) *Polizei* and *Feuerwehr* both covering a public service. *Friday* could be caused by the same phenomenon that caused *Uhr* to appear more frequently in East German newspapers - TV guides, more (cultural) event descriptions, and the like. Likewise, the appearance of *Geburtstag* can be explained by the fact that some newspapers will report on birthdays of (notable or relatively unknown) people in the distribution area. The high appearance of *Bürgermeister* is almost solely caused by the *Nordkurier*; with 445 appearances per million (SD 71.77) it is more than 110 appearances per million higher than the next newspaper - *Thüringer Allgemeine* with 331 per million (SD 47.66). This is again an East German newspaper. This pattern repeats itself

with *Gemeinde*: it appears most in the *Nordkurier* with 334 appearances per million words (SD 75.13), while *Thüringer Allgemeine* takes second place with 228 appearances per million (SD 38.73). A potential explanation for this finding is that these two newspapers report significantly more on local news.
This leaves the two nouns *Verein* and *Gäste.* Both are relatively equally distributed over all East German newspapers; therefore, there might be a synonym for it that is more frequently used in West Germany whereas *Verein* and *Gäste* just appear more in East German newspapers. The online German dictionary Duden (https://www.duden.de/) lists four synonyms for *Verein*: *Bund*, *Gesellschaft, Gemeinschaft,* and *Gruppe*. *Gemeinschaft* did not appear in our data; *Bund* and *Gruppe* both appeared but showed no significant differences between East and West German newspapers (z=.40 and z=-1.46, respectively). *Gesellschaft*, however, appeared significantly more in West German newspapers, z=-2.34. It is plausible that it is used as a synonym, although *Gesellschaft* also carries the meaning of "society" which, according to Duden, is its primary meaning.
For *Gäste*, the most common synonym given by Duden is *Besucher*. According to our data, however, this word is also significantly more common in East German newspapers (z=2.89). *Eingeladene*, another synonym, does not appear in our data. The two nouns *Verein* and *Gäste*, in conclusion, seem to appear more in East German newspapers, and this difference cannot be accounted for by appearance of a synonym. What it does imply is that East German newspapers may tend to publish more localised news - this would be supported by the more frequent appearance of words like *Polizei, Feuerwehr,* and *Bürgermeister*, for example, and maybe also by the more frequent appearance of *Welt* in West German newspapers.
On balance, the results for this method do not seem to imply a lot of lexical differences even though historically, these were the most prevalent linguistic differences between East and West German literature as reviewed in Chapter 2. Most of the lexical differences found seem to be caused by larger amounts of local news only in some newspapers. A surprising number of syntactic differences were found, however. The most notable one is the significantly increased use of commas, even if grammatically incorrect, in West German newspapers; for East German newspapers, the increased appearance of prepositions, partially with case markings, is interesting and puzzling.

Chapter 5: Data-driven approach

5.1 - Method

Lastly, a completely data-driven approach was conducted. This was done for the same data set that was used in the literature-based method; see Chapter 3 for a detailed description of the data set and a justification for inclusion of the chosen newspapers. Compiling a new data set was not necessary, since the goal of this third approach was achievable with the already existing set: the aim was to perform a statistical analysis in order to find possibly existing linguistic differences between East and West German newspapers.

A value was calculated to see if a word or word combination appeared more in East or West German newspapers. This was done as follows. For each feature, the mean frequency of appearance in each Eastern and Western subregion (see fig. 1 for the distributions of the subregions) was calculated, and the two subregions with the highest appearance were taken into account. If the mean frequency of appearance overall was higher in East Germany, the difference between East and West was calculated with the formula

$$x = \frac{mean_{2nd\ place\ subregion\ East} - mean_{1st\ subregion\ West}}{sd_{2nd\ place\ subregion\ East} + sd_{1st\ place\ subregion\ West}}$$

If the mean appearance of the feature in West Germany was higher, the first and second place subregions in the formula were replaced with the respective places from the opposite region. If the value was below 0, a word appeared more in West German newspapers; a value above 0 meant a more frequent appearance in East German newspapers. Only those features with a value above 1 or under -1 were included, to account for random fluctuations in appearance. In addition to this calculation of differences in frequency between the regions, a check was put in place to find words that only appeared in either East or West German papers. The results will be divided into four categories; one each for words that are unique to East and West German newspapers, and one each for words with a score of 1 or higher for East and West. N-grams containing punctuation or proper names will be left out. It stands to reason that proper names of towns, streets, and people are more frequent in the region they are from; additionally, punctuation is likely influenced by the editorial style of the newspaper rather than a regional difference.

5.2 - Results

The results for this method will once again be split, this time into four categories: two for N-grams that were solely found in East and West Germany, and two for N-grams that appeared in both regions but showed a score of 1 or higher for either East or West. At this point, it was decided to only take a closer look at unigrams - the bi- and trigrams found using this method did not add any additional information. Most of them contained words that were already present in the unigrams and in the majority of other cases, the bi- and trigrams contained a substantial amount of noise (quotation marks, other punctuation, etc.). This will be further discussed in the next section.

1 - Only present in West German newspapers

A list of the 25 N-grams taken into account for this category is displayed in table 4. These are the first 25 words from the total list of 14.197 words unique to West German newspapers that were not a proper name (of a town, person, street, or similar) nor an abbreviation - their meanings turned out to be nearly impossible to track down, since some were newspaper-specific and only appeared sporadically.

Many words are specific to traditions, e.g. *Ostereierschießen, Schlussrast* or *Schlusseinkehr*, and *Kappensitzung*. Additionally, a fair few are region-specific names for sports leagues or groups - *Fußball-Gruppenliga* or *Turngesellschaft*, for example. Then, there are some religious nouns (*Pfarrfest, Pfarrgruppe, Kapellchen*) and a fair few buildings specific to towns or cities - *Umwelt-Campus, Stadtteilbücherei, Kinopolis,* or *Naturfreibad*.

Only one word that is not a noun appears in this list - *schwätze.*

Word	**Translation**
Kappensitzung	carnival meeting (wearing weird hats)
Umwelt-Campus	green campus
Pfarrfest	parish festival
Fußball-Gruppenliga	Football group league
Regierungspräsidentin	Female government president
Handball-Bezirksoberliga	upper regional league of a handball competition
Turngesellschaft	Gymnastics club
Traumschleife	Dream loop (nice path to walk on)
Schlussrast	meal at the end of an organised walk
Stadtteilbücherei	library in a city district
Kinopolis	cinema
Ostereierschießen	easter egg shooting
Ultranet	high voltage power grid

Kapellchen	small chapel
Naturfreibad	swimming pool in nature
Pfarrgruppe	parish group
schwätze	chat (1sg, 1/3sg subjunctive)
Bezirksliga-Fußballer	player in regional football league
Ticket-Regional	regional ticket sales point
Familienbildungswerk	(evangelical) center for family education courses
Tumorkranke	cancer patient
Fruchtmarkt	fruit market
Genehmigungsdirektion	approval directorate
Schlusseinkehr	meal at the end of an organised walk
Volleyball-Oberliga	upper regional volleyball league

Table 4: List of the first 25 words only appearing in West German newspapers

2 - Only present in East German newspapers

Again, a list of the 25 N-grams taken into account for this category is displayed in table 5; these are selected in the same way as was done for the first category. The *Landesliga* is a supraregional football league, which in general represents the seventh highest league. There are 19 *Landesligen* in total, and most are further divided into *Landesklassen*. The specific term *Fußball-Landesklasse* was used for six regions in the GDR, which were dissolved in 1952. It seems that compound words with *Landes-* in it are mostly exclusive to East German newspapers.

We also find a few traditions in this list - *Frauentagsfeier, Rentnerfasching,* and *Jugendweihen* -, as well as some words describing (public) buildings: *Milchviehanlage, Schulteil, or Jugendwaldheim,* for instance.

An interesting appearance is made by *Nachholespiel*. Included in the list of words known to be different in East and West Germany (see Appendix A) were *Nachholbedarf* (West) and *Nachholebedarf* (East). It is interesting, then, that another compound word with *Nachhole-* instead of *Nachhol-* is only present in East German newspapers. It should, however, be noted that the word *Nachholspiel* also exists, in far greater frequency than *Nachholespiel*, and shows no significant differences in appearance between East and West German newspapers.

Lastly, there again is only one verb - *beräumen*. It carries the same meaning as *räumen*, which appears in both East and West German newspapers with no significant frequency differences.

Word	**Translation**
Fußball-Landesklasse	football
beräumen	remove

Frauentagsfeier	women's day celebration
Knüppelkuchen	bread made from dough twisted around a stick
Milchviehanlage	dairy cattle facility
Naturparkverwaltung	nature reserve administration
Landesausscheid	supraregional competition
Landesklasse-Aufsteiger	promoted team in supraregional league
Landesklasse-Team	team in supraregional league
Schulteil	part of school
DDR-Liga	league in the GDR
Wahlbereichen	voting regions
Fördermittelbescheide	funding decisions
Rentnerfasching	retirement carnival
Landesklasse-Vertreter	representative of supranational league
Jugendweihen	non-religious celebration at the end of secondary school
Sportlerumfrage	survey among athletes
Landesklasse-Staffel	team from supraregional league
Landesklasse-Kicker	players from supraregional league
Fußball-Landespokals	supraregional cup
Fachkabinette	school rooms specifically for one subject
Landesanglerverband	supraregional fishing association
Nachholespiel	catching up game
Jugendwaldheim	youth home in a forest
Landesklasse-Absteiger	relegated team in supraregional league

Table 5: List of the first 25 words only appearing in East German newspapers

3 - More frequent appearance in West German newspapers

A list of the 25 N-grams taken into account for this category is displayed in table 6. Interestingly, this category only contained 21 words that were not a proper name (street name, first or last name, geographical locations, …) or abbreviation. A large amount of the words present in this category are indications of timespans: seventies, nineties, etc. Also, there are quite a few words related to catholicism. The most striking find is that the period (.) is used significantly more in West German newspapers.

N-gram	Translation	Score
.	.	-3.4477
ihn	him (3sg m acc)	-2.1875
Siebzigerjahren	1970s	-2.1649
Neunzigerjahren	1990s	-1.9063
Er	he	-1.8847
alles	everything	-1.6884
Achtzigerjahren	1980	-1.6531
ihm	him (3sg m dat)	-1.6483
Siebzigerjahre	1970s	-1.4492
Fünfzigerjahren	1950s	-1.4064
lang	long	-1.3804
ein	a	-1.3356
katholischen	catholic	-1.1599
TV	TV	-1.0975
irgendwann	sometime	-1.0835
Info	information	-1.0616
Katholische	catholic	-1.0222
Pfarrzentrum	parish center	-1.0211
Kollegen	colleagues	-1.0150
Sein	his	-1.0110
Katholischen	catholic	-1.0008

Table 6: List of the 25 highest scoring words appearing in newspapers from both parts of Germany, but appearing significantly more in West Germany

4 - More frequent appearance in East German newspapers

A list of the 25 N-grams taken into account for this category is displayed in table 7. Much like in the second category, words that only appeared in East Germany newspapers, a few sports-related words are found here. *beräumen* makes an appearance again, but in conjugated form.

Furthermore, there are quite a few political words and location-related terms, such as *Altkreis* and *Ortsteil*.
Strikingly, several words can be related back to the list of words method one was based on (see Appendix A) - words that were known to be either East or West German. None appear in the exact form in which they were recorded in the past, but nonetheless: *Volkssolidarität* is a compound with *Volk* in it, which used to be typically East German. *Vereinschef* is a compound which contains *chef* - compare with *Parteichef* which used to be typically West German -, whereas there is a synonym *Vereinsvorsitzender*. Lastly and possibly most surprisingly, *Kameraden* makes an appearance in this list.

N-gram	**Translation**	**Score**
parteilos	without a (political) party	3.0245
Fördermittel	funding	2.2326
Kirchgemeinde	parish	2.1814
Landkreis	district	2.0071
DDR-Zeiten	GDR times	1.9485
beräumt	removes (3sg)	1.8389
Einwohner	resident(s)	1.6589
Landesklasse	sports league in a Bundesland	1.6044
Altkreis	former district	1.5980
Kameraden	comrades	1.5900
Eigenmittel	own funds	1.5739
Fördermittelbescheid	funding decision	1.5695
Volkssolidarität	solidarity of the people	1.5676
Vereinschef	club chairman	1.4344
Kraftfahrer	driver of a car	1.3509
Mannschaftsleiter	team leader	1.3361
Fördermitteln	funding	1.3275
Blutprobenentnahme	blood draw	1.2931
Vorhaben	project, intention	1.2920
Orten	place, town	1.2850
siebenten	seventh	1.2711

Ortsteil	district	1.2611
Kirchgemeinden	parishes	1.2298
hiesigen	local	1.2278
informierte	informed (1/2/3sg)	1.2272

Table 7: List of the 25 highest scoring words appearing in newspapers from both parts of Germany, but appearing significantly more in East Germany

5.3 - Discussion

For the third and final method of this research, a data-driven approach was taken to find potential differences in the frequency of appearance of N-grams between West and East German newspapers. In addition to looking for words that solely appear in either East or West German newspapers, a list was compiled of words that scored higher than 1 (substantially more present in East German newspapers) or lower than -1 (higher appearance in West German newspapers). For a detailed explanation of how these scores were calculated, please refer back to section 5.1. The results were divided into four categories; words exclusively appearing in West German newspapers, words exclusively appearing in East German newspapers, ones that appear significantly more in West German newspapers, and finally ones that appear significantly more in East German newspapers. For the full 25-word lists for each of these categories, please refer back to tables 4 through 7 in the previous section.

As mentioned in that section, only unigrams were taken into account. After looking at all N-grams, it was noted that including bi- and trigrams did not add new information - e.g., where the unigrams already included *Fünfzigerjahren*, the bi- and trigrams would add *in den Fünfzigerjahren, den Fünfzigerjahren* and so on. If we take the category of words appearing significantly more in West German newspapers, this only really meant leaving out eight bi- or trigrams with new information. These are listed in table 8 below:

N-gram	**Translation**	**Score**
dieser Stelle schreiben	write [at] this point	-2.8997
Weitere Auskünfte	Further information	-1.1828
Teilnahme ist kostenlos	Participation is free of charge	-1.1737
Ein Gespräch	A conversation	-1.1463
Anmeldung bei	Registration with	-1.0435
Schreiben Sie eine	(*polite*) write a	-1.0430
dieser Seite	[on] this page	-1.0430
man ihm	someone (...) him	-1.0045

Table 8: List of bi- and trigrams excluded for the category of words appearing significantly more in West German newspapers

All of these bar *man ihm* and *Ein Gespräch* can be interpreted as indications of a newspaper-specific trait: a competition readers can participate in, invitations to write feedback, a notice to go to the internet for more information. The data support the assumption that these are newspaper-specific: the highest-scoring *dieser Stelle schreiben*, for instance, only appears in four out of 19 newspapers with frequencies between 0.04 per million and 3.27 per million, except in *Der Spiegel* where it appears with a frequency of 23.98 per million words. As researching these very local quirks was not the goal of this paper, they are not relevant and there is no harm in discarding them. *Ein Gespräch* follows the same pattern, showing it is newspaper-specific: it appears in all newspapers, generally with a frequency between 0 and 2 per million words but never exceeding 9 per million, except in *Die Zeit*, where it appears with a frequency of 32 per

million words. This does, however, leave *man ihm* - this is new information not covered by the unigrams, it is not noise (metadata, punctuation etc.), and can be an indication of a specific syntactic construction with a conditional verb, e.g. (...) *hätte man ihm* (...). However, since it could also appear in a construction with a normal verb (*hat man ihm*) or something completely different, more context would be needed than can be covered by the dataset for this current research. Moreover, it is only marginally above our criterion with a score of -1.0045 - it should not be weighed too heavily as a difference between East and West German newspapers. For the other categories, similar patterns were revealed; therefore, with the same reasoning as used above, it is assumed that by removing bi- and trigrams, little to no significant new information is lost.

The most striking difference between the results for this method and the results for the other methods is the presence of so many nouns. The two categories of words that exclusively appear in either region, for instance, both contain only one non-noun (it is a verb in both). For the purposes of this research, however, the other two categories would yield more relevant results. Taking the words only appearing in East German newspapers as an example, all appear with a frequency at or below one per million words, except the top two (*Fußball-Landesklasse* at 6 per million and *beräumen* at 2 per million). By comparison, quite a few of the nouns in the literature-based method appeared between 1 and 10 times per million words, with a significant number of them appearing between 10 and 100 times per million. In this research-based method, *Politik* appeared between 86 and 384 times per million words, depending on the newspaper. In general, the nouns found using our third method are not very frequent, so it should be kept in mind that they generally do not represent a set of words that absolutely dominates the newspapers in either East or West. Additionally, in most cases the standard deviation for a newspaper is usually greater than the mean appearance, indicating that usage of those words is highly volatile over time.

However, looking at the words exclusively appearing in either East or West German newspapers, a pattern is clearly visible. The vast majority of exclusively East German words are sports terms, specifically words containing the term *Landesklasse*. On the other hand, while not as overwhelming, there are several exclusively West German words containing *-liga*. This is a remnant of GDR-times: the football system in East Germany used to consist of six *Landesklassen*, whereas West Germany had a similar tier called *Landesliga*. The *Landesklassen* were dissolved in 1952, when the six *Bundesländer* as they are known today were formed, and replaced with six *Landesligen*. Even today, 70 years and several system changes later, some *Landesligen* still get referred to as *Landesklassen* - the website fussball.de, owned by the national football association DFB, still shows several *Landesklassen* for Sachsen, a former GDR *Bundesland*. Even though the term should be obsolete, its continued usage seems to be an artefact of GDR-times.

A lot of regional and local words appear in the lists of words in Chapter 4; mainly in the categories featuring exclusively East or West German words, but also in the other two. Take, for instance, West German-only *Kappensitzung, Pfarrfest,* or *Ostereierschießen*, and East German-only *Frauentagsfeier, Knüppelkuchen,* or *Rentnerfasching.* All are local traditions. Looking at the other two categories, a (local) theme of catholicism appears more in West German newspapers - this is not surprising, considering the West and South are traditionally catholic and the north-east protestant; see fig. 27 for an overview of catholicism in Germany as of 2020 (*image omitted due to copyright issues*). Protestantism does not appear as clearly and as often in East German newspapers, but still makes an appearance in the form of *Kirchgemeinde* - a term used in the evangelical church to describe a local parish.

Fig. 27: Map of Germany showing the percentage of inhabitants that were registered as Catholic in 2020

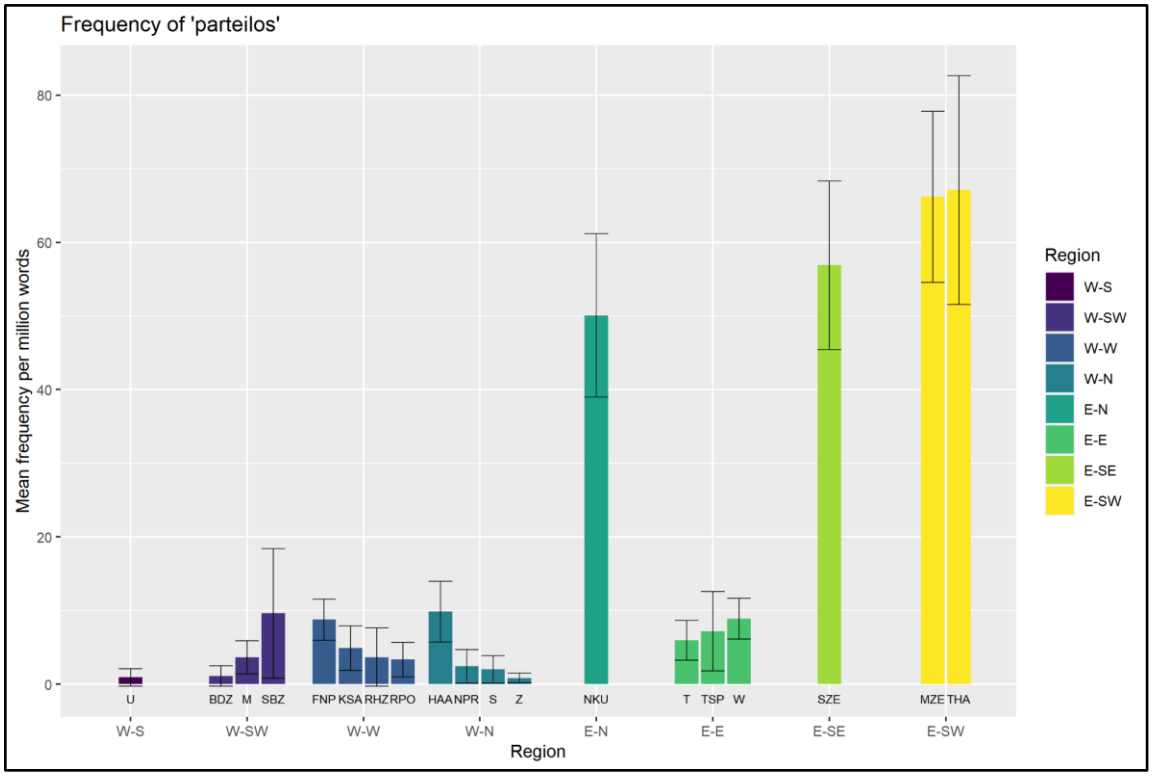

Fig. 28: Mean frequency of appearance per million words of parteilos

Like with the previous method, the frequency of appearance of some words is inflated by just one newspaper. Even though this method attempts to control for that by not looking at the single highest appearance in any newspaper, but at the highest appearance in two subregions, it seems to fail at some points. Take, for instance, the word with the highest score in the category of more frequent appearance in East German newspapers: *parteilos*. Left out of the results section but still appearing in our data are the combinations *parteilos), (parteilos,* and *(parteilos)*. This word, therefore, is a result of newspapers adding information to a politician's name when it is mentioned. Assuming there is not a disproportionate amount of partyless politicians in East Germany, this is merely the result of some newspapers consistently writing this information down whereas others do not. This is supported by graphing the data, as seen in fig. 28. Similarly, a lot of decades are mentioned more frequently in West German newspapers - this can either be a stylistic difference where West German newspapers like to write it out more (*Siebzigerjahre*) whereas East German newspapers abbreviate (*70er Jahre*), or it can be a difference in topics where West German newspapers tend to write more historical articles. Our data does not give any clues as to which of the two would cause this pattern, but it is safe to assume either way that this is not a major lexical difference between East and West Germany.
Comparing the results found using this method with the results for the other two methods uncovers some interesting things. In the research-based method, the comma turned out to be the most significant N-gram that appeared more in West German newspapers by far. Here, this is the period. This is seemingly contradicting - frequent usage of the comma points to a tendency to make longer sentences, whereas frequent appearance of the period suggests shorter sentences. Therefore, it seems that this is a result of the method used - going off the statistical method used by Hellmann (1984) yields different results than going off the two subregions with the highest mean frequency of appearance does. The results for both the research-based and data-driven methods are visualised in figs. 29 and 30. In both images, an X above the column

denotes newspapers that have been looked at for the research-based method; for the data-driven method, all newspapers were taken into account. As a result of the specific newspapers chosen in the research-based method, the comma seemed to be very significantly more common in West German newspapers. Looking at all newspapers, however, shows that this pattern is merely caused by the chosen newspapers - some of the newspapers with the highest frequency of the comma in the West and the lowest frequency in the East were chosen. The higher appearance of the period in the nineteen newspapers looked at for the data-driven method, then, seems to be influenced mainly by the *Mannheimer Morgen* - but even when leaving that newspaper out, its appearance is still well below the mean in all but one East German newspaper. This result seems to be robust, whereas the higher appearance of the comma was an artefact caused by the specific method.

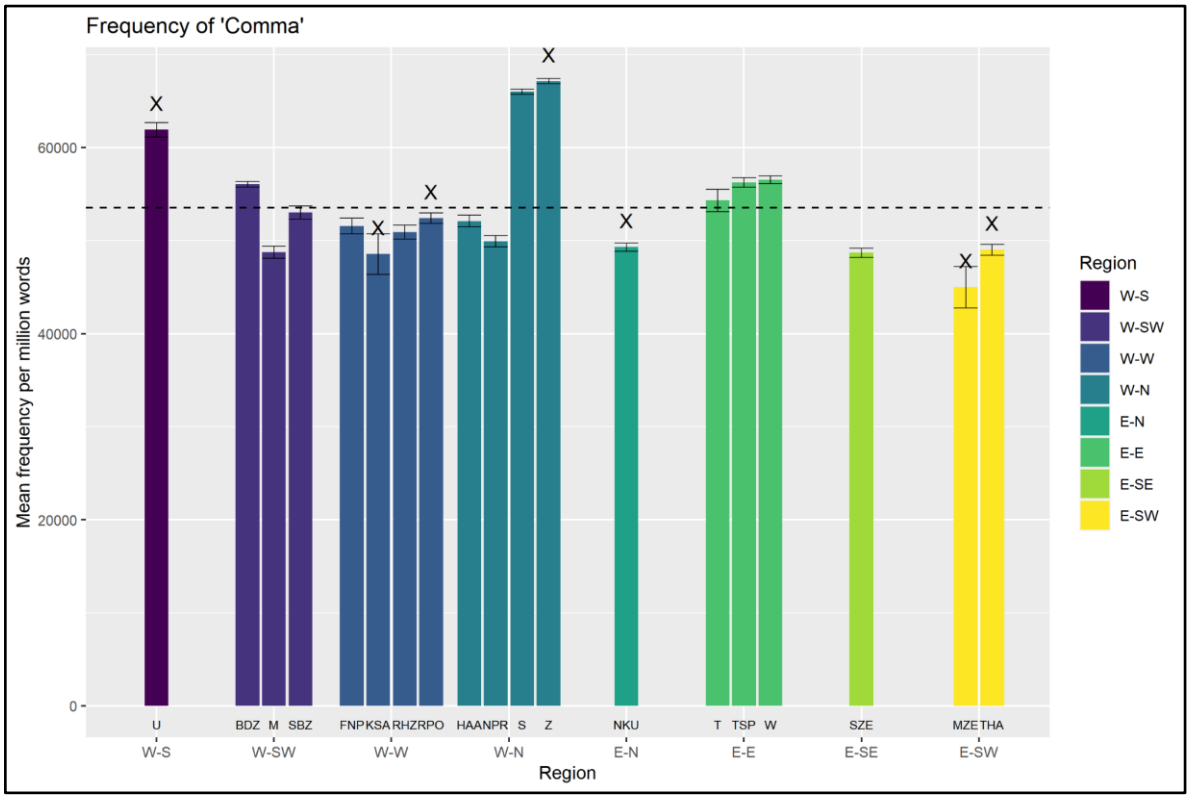

Fig. 29: Mean frequency of appearance per million words of the comma, with newspapers used in the dataset for the research-based method denoted by an X above the column

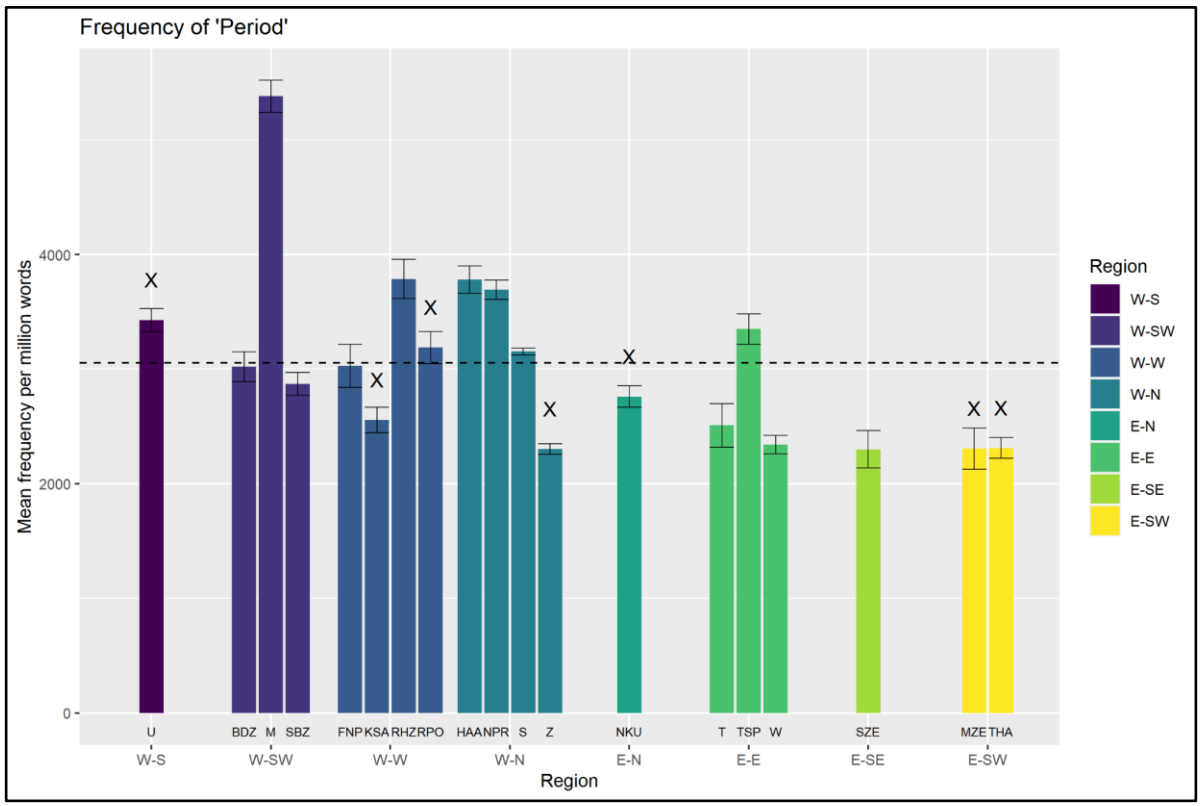

Fig. 30: Mean frequency of appearance per million words of the period, with newspapers used in the dataset for the research-based method denoted by an X above the column

As mentioned in section 5.2 already, the word *Kameraden* is used significantly more in East German newspapers. This is, of course, traditionally a GDR word; much like some of the results of the literature-based approach, however, it is hard to say if this number is artificially inflated by popular scientific articles about traditional GDR words. When compared to the results of the literature-based approach, some other words jump out - mostly words appearing more in East Germany. Take *Volkssolidarität* - *Volk* and compounds containing it were traditionally found to be more frequent in East Germany (Hellmann, 2008a). Note, however, that the results for the literature-based approach indicated that *Volk* was now significantly more common in West German newspapers. The more frequent East German appearance of *Vereinschef* is also quite interesting - *Verein* was also found to be more frequent in East German newspapers in the results for the research-based approach of method 2, but *-chef* as part of the word *Personalchef* used to be more frequent in West Germany, as shown in the literature-based approach. Lastly, the word *Nachholespiel* only appears in East German newspapers, which again can be related back to the list of words used for the literature-based approach (see Appendix A): traditionally, *Nachholbedarf* appeared more in West German literature whereas *Nachholebedarf* with its extra connective -e- appeared more in East German literature. To check if this pattern appears here too, *Nachholspiel* was looked at. It also appears in the data, but has a score of -0.0897 - appearing slightly more in West German newspapers, but not even close to our criterion threshold of -1. It should be noted, however, that *Nachholespiel* only appeared in four newspapers with a frequency of between 0.22 and 1.88 per million and a standard deviation that was higher than the mean in all four cases. *Nachholspiel,* on the other hand, appeared in 17 of 19 newspapers and with a frequency of between 0.06 and 13.8 per million and a standard deviation lower than the mean in roughly half of the cases. *Nachholespiel*, therefore, seems to be a very infrequently and, over time, very inconsistently used word for which there is a more

generally used and accepted replacement, *Nachholspiel*. This is a pattern we have seen in the results for the literature-based approach as well, e.g. with *Kaufhalle,* which still appears more in East Germany, but a lot less frequently than its synonym *Supermarkt*.
Overall, the results for this data-driven approach do not give any indication that lexical differences between East and West Germany today are as significant and widespread as they were decades ago. Some historic patterns and words, like *Kameraden, Nachholespiel,* or names for sports leagues, do show up; they do not, however, indicate systematic differences but rather infrequently used remnants.

Chapter 6: Conclusion

6.1 - Study limitations and recommendations

The findings of this study should be seen in light of some limitations. Most notably, the dataset obtained with the help of the *Institut für Deutsche Sprache* was limited: because of contracts with the newspapers included in DeReKo, we were not allowed to see words in their full context. Individual words and even sentences were allowed, but never the full article. Had there been access to the full articles, the results of this study could have been placed in that light. For instance, a more conclusive answer could have been found for the question if the results for the literature-based method - some traditionally East German words like *Broiler* still being more common in East German newspapers - are at least partially caused by popularising news articles about their appearance, rather than actual use of the words. Additionally, the *Gebrauchsspezifika* and *Wertungsspezifika* mentioned in Chapter 2 could have conceivably been looked at; maybe not in as much detail since these would have had to be vetted by hand, but they could have been included nonetheless.

Whereas lexical and syntactic differences could be, and were, looked at during this study, semantics had to be left out based on time constraints and, in some cases, lack of material. Studying semantics benefits greatly from having the context material available - provided it can be obtained as per the recommendation above, things like word sense disambiguation could be applied to the dataset. In the literature-based method, for instance, the word *Preis* was looked at. Historically, this word in the sense of "cost" appeared more in West German literature; however, it can also mean "prize". In the current study, there was no way of distinguishing between which of the two senses it carried, but if the context is given, new software could be written to try and differentiate between senses. Furthermore, DeReKo itself is tagged for some features like named entities; these tags will help with extracting more information from the dataset, but were not taken into account in the current study. It is recommended that this be done in a future study.

In the results for the literature-based method, the findings for Berlin-based newspapers stood out. Quite frequently, they would behave more like West German newspapers did (see for instance figs. 10, 11, and 12). This in itself is not very surprising; Berlin is a metropolis with a relatively diverse demography. According to the *Amt für Statistik Berlin-Brandenburg*, approximately 3.6 million people lived in Berlin in 2020 and of those, 717.600 (19.5%) were foreign nationals. The *Statistisches Bundesamt* keeps track of the same numbers for the entire country; they recorded 83 million people living in Germany in 2021, with 10.8 million (13%) being foreign nationals. This comparatively large portion of foreign nationals in Berlin might influence the day-to-day language: it might be more internationally-inclined with more English words, simpler sentences, or less regional language. The findings of this study suggest that Berlin newspapers tend to align more with West German newspaper language, which historically was more English-oriented, as discussed in Chapter 2. Hellmann (1984) focused on readership as a basis for selecting his newspapers. However, it is the language of the authors which is being measured rather than that of the readers. This means we assume that the language use of authors and readership is somehow linked, probably by way of living in the same area. In Hellmann's time, this may have been the case, but nowadays this link is possibly weaker. Especially in Berlin, we can expect that many journalists moved there, for instance to move closer to the capital, bringing an influence from the West; this is assuming that journalists writing for Berlin newspapers actually live in Berlin. Checking this hypothesis can, however, not

be done with the data we have at our disposal. It would be necessary to map which authors wrote what articles, and what their background is. Additionally, because the pattern of Berlin newspapers behaving differently to East German newspapers was found with only a few of the 77 words looked at, a further study that focuses specifically on Berlin-based newspapers would be necessary to explore how robust this pattern is and to look into the finer workings of it.
The research-based method was a close replication of Hellmann (1984). While the method and statistics themselves were not hard to copy since they were well-documented, today's newspapers do not represent the population as accurately as newspapers in the 1980s did. This is in part because newspapers are, or claim to be, politically independent now, but also because overall readership has declined steeply. *Neues Deutschland* for example, a newspaper Hellmann looked at as it was the leading GDR newspaper at the time, had a circulation of 1.1 million newspapers in 1989, as reported by the newspaper itself (https://www.nd-aktuell.de/kontakt/9). In the fourth quarter of 2021, this had declined to a circulation of 17.612 according to the IVW (2021), a German institute keeping track of circulation of countless newspapers, magazines, brochures, and the like. Other newspapers showed a similar decline; thus, a single newspaper cannot hope to be representative of a significant part of the population anymore. Therefore, we had to try and approximate the coverage of Hellmann's (1984) newspapers: in that study, newspapers were chosen based on how well their readership represented the general population, rather than based on their language - as said before, language in newspapers represents the language of their writers rather than their readers. Seeing as the coverage of all newspapers shared the same fate, even selecting 50 newspapers might not have covered the entire population; therefore, we took newspapers that covered a widespread distribution area and for which there was a large amount of data present in DeReKo, with the assumption that they would cover as much of the readership population as possible. However, this was impossible to prove conclusively, and might have introduced a bias in the data. Unfortunately, due to strict privacy laws and because the companies themselves don't always make collection of a large amount of data possible, this was not feasible for this study but would be a recommendation for future research. Additionally, it may be desirable to look into possible differences between language of the readers and language of the writers - even though the research-based method was a close copy of Hellmann (1984) and therefore copied its method as well, it is not the language of the readership that is being measured, but rather the language of the writers. The results cannot be generalised to the entire population if there are grave differences in language between readers and writers; while this did not fall within the scope of the current research, it is recommended to look into this in a future study.
A final recommendation is based on the fact that the dataset consists of data from only one year, 2019. As such, a diachronic study would be desired; it would, for instance, give more information about the stability of the word use over time. Hellmann (1984) built this into their design by using a dataset that was compiled over 25 years: the data included newspapers from 1949 up until 1974, in intervals of 5 years. Given the contents of DeReKo (not all years are present and accessible for all newspapers), it was not feasible to include data from such a timeframe in the current study. However, adding data from at least another year should yield valuable additional information regarding the stability of the usage of N-grams over time, so that conclusions could also be drawn regarding whether some N-grams are on the way out, being used less and less, or if they are a stable but infrequent part of the German language.

6.2 - Conclusion

Overall, the findings of the literature-based method indicate that some traditionally East or West German words are still more prevalent in the area they were historically found more in. This mostly happens for traditionally East German N-grams: words like *Broiler* or *Kaufhalle* do still show up more in East German newspapers, but in some cases (like *Kaufhalle*), this excludes Berlin-based newspapers. Several words that were historically more prevalent in either half of the country and were once defining linguistic differences between the two parts of Germany, however, were very infrequently or not at all used (12 out of 77 N-grams) in the newspapers included in the dataset. The majority of the remaining words, 47 out of 77, appear with a low frequency - more than 0 times but less than 10 times per million words. Even though some of these do display differences in frequency of appearance between East and West Germany, it can hardly be said they constitute a fundamental linguistic difference anymore; after all, they only make up a very small percentage of all the words used in newspapers today.

The results for the research-based method did not turn up many lexical differences at all; even though in Chapter 2, it was discussed that these constituted the majority of linguistic differences between East and West German literature, the N-grams differing most significantly in frequency of appearance between East and West German newspapers mostly indicated syntactic differences. For instance, the comma was the N-gram that most significantly appeared more in West German newspapers; the top 40 included twelve more N-grams consisting of a comma and a pronoun or conjunction. This, however, seems to have been caused by a bias in the dataset - the selected West German newspapers just so happened to have an unusually high amount of commas, whereas the East German newspapers had an unusually low amount of them. These results, therefore, could not be generalised; it is recommended to conduct a study with a different, more representative dataset in order to be able to draw conclusions about linguistic differences in comparison with the study compared to in this method, Hellmann (1984).

The findings of the last method, a data-driven statistical analysis, indicated that many words either only appearing in one half of Germany or appearing significantly more in one of the parts were nouns. Again, most of them did not appear with a high frequency (usually around once per million words), and the majority could be chalked up to either newspaper-specific habits (like writing down *parteilos* or consistently talking about events in past decades, for instance in a specific history section) or regional differences (like religious words). A select few seem to show lexical differences that were also attested to in historic literature: *Nachholespiel* appears more in East German newspapers when its counterpart *Nachholspiel* does not show this difference, and *Kameraden* still appeared more in East German newspapers as well. However, this is such a small part that again, this does not seem to indicate any fundamental differences between East and West German newspapers, but rather remnants of the past.

A few additional general recommendations are made with regards to future research: provided that more contextual data is acquired, it would be beneficial to take semantics into account. Word sense disambiguation could not be done for the current study, but would provide valuable insight into usage of different senses of the same word, for instance. Additionally, the corpus DeReKo is already tagged for certain features like named entities. Due to time constraints, many could not be taken into consideration for the current study; it would be recommended to look into this in a future study. Lastly, repetition of this study for a dataset from a different year in order to perform a diachronic study is recommended to view word usage over time; after all, the data included were only from one year and as such, only provide a snapshot.

References

Amt für Statistik Berlin-Brandenburg (2020). https://www.statistik-berlin-brandenburg.de/bevoelkerung/demografie/bevoelkerungsstand. Retrieved Jan 22nd, 2022.

Antos, G., & Schubert, T. (1997). Unterschiede in kommunikativen Mustern zwischen Ost und West. *Zeitschrift Für Germanistische Linguistik, 25*(3), pp. 308-330.

Bock, R., Harnisch, H., Langner, H., & Starke, G. (1973). Zur deutschen Gegenwartssprache in der DDR und in der BRD. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, 26*(5), pp. 511-532.

Dieckmann, W. (1967). Kritische Bemerkungen zum sprachlichen Ost-West-Problem. *Zeitschrift der deutschen Sprache, 23*, pp. 136-165.

Folsom, M., & Rencher, A. (1977). Zur Frage der sprachlichen Unterschiede in der BRD und der DDR: Zwei statistische Studien. *Deutsche Sprache, 1*, pp. 463-468.

Hellmann, M. (1976). *Bibliographie zum öffentlichten Sprachgebrauch in der Bundesrepublik Deutschland und in der DDR.* Düsseldorf, Germany: Pädagogischer Verlag Schwann.

Hellmann, M. (1980). Deutsche Sprache in der Bundesrepublik Deutschland und der Deutschen Demokratischen Republik. In H. Althaus, H. Henne, Helmut, H. Wiegand (Eds.): *Lexikon der Germanistischen Linguistik* (3). Tübingen, Germany: Niemeyer.

Hellmann, M. (1984). *Ost-West-Wortschatzvergleiche. Maschinell gestützte Untersuchungen zum Vokabular von Zeitungstexten aus der BRD und der DDR.* Tübingen, Germany: Narr.

Hellmann, M. (1990). DDR-Sprachgebrauch nach der Wende - eine erste Bestandsaufnahme. *Muttersprache 100*(2-3), pp. 266-286.

Hellmann, M. (2008a). *Sprache zwischen Ost und West - Überlegungen zur Wortschatzdifferenzierung zwischen BRD und DDR und ihren Folgen.* Hildesheim a.o., Germany: Olms.

Hellmann, M. (2008b). Zur Sprache vor und nach der „Wende" – Ost-West-Kulturen in der Kommunikation. In H. Casper-Hehne & I. Schweiger (Eds.), *Deutschland und die "Wende" in Literatur, Sprache und Medien* (pp. 97-116). Göttingen, Germany: Universitätsverlag Göttingen.

IVW (2021). https://www.ivw.de/aw/print/qa/titel/1680. Retrieved Jan 22nd, 2022.

Jarausch, K. (2012). Historische Texte der DDR aus der Perspektive des "linguistic turn" [1998]. *Historical Social Research, 24*, pp. 229-248.

Katholische Kirche in Deutschland (2021). *Zahlen und Fakten.* *https://www.dbk.de/fileadmin/redaktion/Zahlen%20und%20Fakten/Kirchliche%20Statistik/Allgemein_-_Zahlen_und_Fakten/AH-325_DBK_BRO_ZuF_2020-2021_Ansicht.pdf*

Kennetz, K. (2010). German and German Political Disunity: An Investigation into the Cognitive Patterns and Perceptions of Language in Post-Unified Germany. In *Perceptual Dialectology*. Berlin, Germany/ Boston, MA: De Gruyter.

Kreutz, H. (1997). Sprachliche Wiedervereinigung Ost-West. Eine pragmalinguistische Untersuchung zu Erscheinungen kommunikativer Unsicherheit bei jungen Ostbürgern. *Institut für Deutsche Sprache, 1*(2), pp. 1-317.

Lerchner, G. (1992). Broiler, Plast(e) und Datsche machen noch nicht den Unterschied. Fremdheid und Toleranz in einer polyzentrischen deutschen Kommunikationskultur. In Gotthard Lerchner (ed.) *Sprachkultur in Wandel. Anmerkungen zur Kommunikationskultur.* Frankfurt, Germany: pp. 297 - 332.

Maas, U. (2012). Was ist Deutsch? Die Entwicklung der sprachlichen Verhältnisse in Deutschland. Unter Mitarbeit von Solvejg Schulz, *Zeitschrift für Rezensionen zur germanistischen Sprachwissenschaft, 5*(2), pp. 196-201.

Markkanen, R. & Schröder, H. (Eds.) (1996). *Hedging and Discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts.* Berlin, Germany: De Gruyter.

Moser, H. (1954). Entwicklungstendenzen im heutigen Deutsch. *Der Deutschunterricht, 6*(2), pp. 87-107.

Pankanin, E. (2017). Die ostdeutsche Realität abgebildet in der Sprache. Ist Ostalgie nach wie vor präsent im heutigen Deutsch? *Heteroglossia, 7,* pp. 107-122.

Plewnia, A., & Rothe, A. (2009). Eine Sprach-Mauer in den Köpfen? Über aktuelle Spracheinstellungen in Ost und West. *Deutsche Sprache, 37*(3), pp. 235-279.

Polenz, P. von (1999). *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart, Band 3*. Boston, MA: De Gruyter.

Polenz, P. von (1993). Die Sprachrevolte in der DDR im Herbst 1989. *Zeitschrift für germanistische Linguistik, 21*(2), 127-149.

Reich, H. (1968). *Sprache und Politik. Untersuchungen zu Wortschatz und Wortwahl des offiziellen Sprachgebrauchs in der DDR.* München, Germany: Max Huber Verlag.

Schmidt, H. (2007). Auferstanden aus Ruinen - Sprachliche Erbstücke aus schwierigen Zeiten. *Sprachreport 3*, pp. 2-11.

Schmidt, H. (2009). Über den gemeinsamen Sprachgebrauch in Ost und West, seine Problemen und kreativen Möglichkeiten. *Deutsche Sprache, 37*(3), pp. 97-129.

Schlobinski, pp. (2015). Das Berlinische in der Einschätzung der Bürger der Hauptstadt. *Muttersprache, 1*, pp. 2-13.

Schlosser, H. (1991). Deutsch in Ost und West. *Der Sprachdienst, 15*(1), pp. 26-31.

Schlosser H.D. (2001) Deutsche Teilung, deutsche Einheit und die Sprache der Deutschen. In I. Kühn (Ed.) *Ost-West-Sprachgebrauch — zehn Jahre nach der Wende.* Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Statistisches Bundesamt (2021). https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/liste-zensus-geschlecht-staatsangehoerigkeit.html. Retrieved Jan 22nd, 2022.

Steffens, D. (2009). 20 Jahre Mauerfall - Zur Wortschatzentwicklung seit der Wendezeit. *Deutsche Sprache, 37*(3), pp. 148-167.

Stickel, G. (2000). Was West- und Ostdeutsche sprachlich voneinander halten. In R. Reiher, A. Baumann (Eds.) *Mit gespaltener Zunge? Die deutsche Sprache nach dem Fall der Mauer.* Berlin, Germany: Aufbau-Taschenbuch-Verlag.

Uchimura, K. (1983). Zur Entwicklung der deutschen Sprache in der DDR. *Studies in the Humanities, 10,* pp. 93-105.

Young, J. (2005). From LTI to LQI: Victor Klemperer on Totalitarian Language. *German Studies Review, 28*(1), pp. 45-64.

Appendix A: List of words used for literature-based approach

1 - *Bezeichnungsspezifika*
Categories:
I - not present in dataset
II - 0<□< 10 per million
III - 10<□<100 per million
IV - >100 per million

West	East	Translation	Source	Category
Plastik	*Plast(e)*	plastic	Hellmann 1980/1984	West: III East: II
Staatsangehörigkeit	*Staatsbürgerschaft*	nationality	Hellmann 1980/1984	West: II East: II
Arbeitnehmer	*Werktätiger*	employee	Hellmann 1980/1984	West: III East: I
Personalchef	*Kaderleiter*	personnel manager	Hellmann 2008a	West: II East: I
Mähdrescher[1]	*Kombine*	combine harvester	Hellmann 2008a	West: II East: I
Brathähnchen[2]	*Broiler*	roast chicken	Hellmann 2008a	Both: II

[1] Used in both East and West Germany
[2] Used in both East and West Germany

Aerobik	*Popgymnastik*	aerobics	Hellmann 2008b	Both: I
Diskjockey	*Schallplatten-unterhalter*	DJ	Hellmann 2008b	Both: II
Fruchtsaft	*Juice*	fruit juice	Hellmann 2008b	Both: II
Flüchtling/ Heimatvertriebener	*Umsiedler*	refugee	Hellmann 2008b	West: II/I East: II
Nachholbedarf	*Nachholebedarf*	need to catch up	Hellmann 2008b	Both: II
regional	*territorial*	regional	Hellmann 2008b	Both: II
Supermarkt	*Kaufhalle*	supermarket	Hellmann 2008b	West: III East: II
Zielsetzung	*Zielstellung*	goal, objective	Hellmann 2008b	Both: II
Kita	*Krippe*	daycare	Schlobinski (2015)	West: III East: II

shopping/shoppen	*einkaufen*	shopping	Stickel (2000)	Both: II
Flieger	*Flugzeug*	plane	Stickel (2000)	West: II East: III
ich	*ick/icke*	I	Stickel (2000)	West: IV East: II
ja	*nu/nü/no*	yes	Stickel (2000)	West: IV East: II/I/II

2 - *Häufigkeits- und Lexemspezifika*

West	**Translation**	**Source**	**Category**
freiheitlich	liberal	Hellmann 1980/1984	II
Partnerschaft	partnership	Hellmann 1980/1984	III
dynamisch	dynamic	Hellmann 1980/1984	II
Markt	market	Hellmann 2008a	III
Demokratisierung	democratisation	Hellmann 2008a	II

Preis	cost	Hellmann 2008a	III
demokratischer Sozialismus	democratic socialism	Hellmann 2008a	I
europäische Integration	European integration	2008a	II
konzentrierter Aktion	concentrated action	Hellmann 2008a	I
Prager Frühling	Prague spring	Hellmann 2008a	II
Super	great	Stickel 2000	III
Ossi(s)	people from East Germany	Stickel 2000	Both: II
Kids	children	Stickel 2000	II
okay/o.k.	okay	Stickel 2000	*okay:* III *o.k.*: I

East	**Translation**	**Source**	**Category**
sozialistisch	socialist	Hellmann 1980/1984	II
Produktion	production	Hellmann 1980/1984	III

wir/unser	us/our	Hellmann 1980/1984	IV
schöpferisch	creative	Hellmann 1980/1984	II
allseitig	universal	Hellmann 2008a	II
friedliebend	peace loving	Hellmann 2008a	II
Volk[3]	people, nation	Hellmann 2008a	III
Massen	masses	Hellmann 2008a	II
Qualifizierung	qualification	Hellmann 2008a	II
friedliche Koexistenz	peaceful coexistence	Hellmann 2008a	II
umfassender Aufbau	comprehensive structure	Hellmann 2008a	I
wissenschaftlich-technische Revolution	scientific-technical revolution	Hellmann 2008a	I
ca.	approx.	Kreutz 1997	II

[3] Includes compounds containing this word

etwa	approximately	Kreutz 1997	IV
annähernd	nearly	Kreutz 1997	II
vielleicht	maybe	Kreutz 1997	IV